(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
8 March 2007 (08.03.2007)

PCT

(10) International Publication Number
WO 2007/027194 A2

(51) International Patent Classification:
G01N 21/65 (2006.01) *G01N 21/55* (2006.01)

(21) International Application Number:
PCT/US2006/006618

(22) International Filing Date:
23 February 2006 (23.02.2006)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
11/068,739 28 February 2005 (28.02.2005) US

(71) Applicant *(for all designated States except US)*: **THE BOARD OF TRUSTEES OF THE UNIVERSITY OF ILLINOIS** [US/US]; 352 Henry Administration Building, 506 Wright, Urbana, IL 61801 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **MARKS, Daniel, L.** [US/US]; 820 E. Kerr Avenue, Urbana, IL 61801 (US). **BOPPART, Stephen, A.** [US/US]; 4306 Stonebridge Court, Champaign, IL 61822-9345 (US).

(74) Agent: **RAUCH, Paul, E.**; EVAN LAW GROUP, LLC, 566 West Adams, Suite 350, Chicago, IL 60661 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

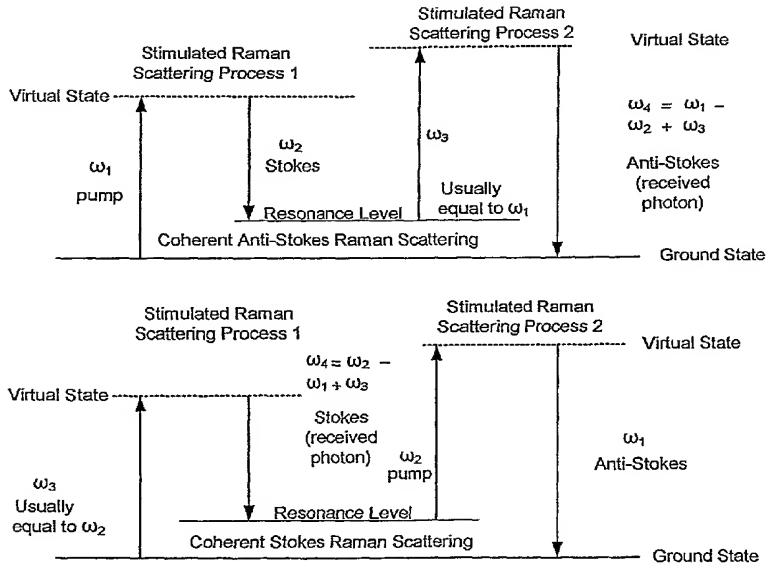
(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: DISTINGUISHING NON-RESONANT FOUR-WAVE-MIXING NOISE IN COHERENT STOKES AND ANTI-STOKES RAMAN SCATTERING



(57) Abstract: A method of examining a sample comprises exposing the sample to a pump pulse of electromagnetic radiation for a first period of time, exposing the sample to a stimulant pulse of electromagnetic radiation for a second period of time which overlaps in time with at least a portion of the first exposing, to produce a signal pulse of electromagnetic radiation for a third period of time, and interfering the signal pulse with a reference pulse of electromagnetic radiation, to determine which portions of the signal pulse were produced during the exposing of the sample to the stimulant pulse. The first and third periods of time are each greater than the second period of time.

WO 2007/027194 A2

DISTINGUISHING NON-RESONANT FOUR-WAVE-MIXING NOISE IN COHERENT STOKES AND ANTI-STOKES RAMAN SCATTERING

FEDERALLY SPONSORED RESEARCH OR DEVELOPMENT

[01] The subject matter of this application may have been funded in part under Contract No. NAS2-02057 awarded by the National Aeronautics and Space Administration (NASA). The U.S. Government may have rights in this invention.

BACKGROUND

[02] Molecules frequently have molecular resonance frequencies that are due to the electromagnetic attractions of atoms in the molecule. These frequencies are those of molecular vibrations, molecular rotational motions, the excitation of electrons to higher energy states, and occasionally finer structures such as hyperfine interactions and optical-magnetic properties. These properties are present without the introduction of any external contrast molecule. These frequencies are usually in the mid-infrared, corresponding to photons of 1.5-50 microns of wavelength. Unfortunately, they cannot be directly excited by electromagnetic radiation of the same frequency because when they are in tissue, the surrounding water absorbs almost all of these frequencies. The range of wavelengths that the tissue is relatively transparent to is 0.6-1.5 microns. Therefore multiphoton nonlinear processes need to be employed to probe these resonances. The photons to stimulate and record the processes are typically in a region where the tissue is not absorbing, so that they can reach the tissue feature and be measured from the feature.

[03] Raman spectroscopy, first discovered in 1928, uses molecular resonance features of frequency $\Delta\omega$ to split a photon of frequency ω into another photon of frequency $\omega - \Delta\omega$ and a resonance excitation of frequency $\Delta\omega$. The presence of photons at frequency $\omega - \Delta\omega$ identifies the concentration of the resonance feature. This process is in practice very weak and requires large amounts of power to produce any detectable amount of photons. This weakness is due to the fact that the probability of a Raman excitation process to occur is proportional to the number of photons at frequency

$\omega - \Delta\omega$ already present, of which there are typically few or none. Since photons that would be emitted by Raman excitation at frequency $\omega - \Delta\omega$ are indistinguishable from the incoming radiation that stimulates them, this is not a viable technique for achieving molecular sensitivity.

[04] Coherent Anti-Stokes Raman Scattering (CARS) is another nonlinear spectroscopy technique that unlike conventional Raman spectroscopy, allows all of the photons necessary to stimulate the process to be introduced into the tissue by the illuminating source. This enables the probability of a CARS interaction to be increased to a (theoretically arbitrarily) high level so that a sufficient number of photons can be produced as to enable detection within a reasonable time period. It is essentially two stimulated Raman scattering processes in parallel. Two photons, a "pump" of frequency ω_1 and a "Stokes" of frequency ω_2 illuminate the tissue. They must be separated in frequency by $\omega_1 - \omega_2 = \Delta\omega$, which is the frequency of the molecular resonance. When molecules of the target molecular species are present, the resonance will be excited, and the pump photon will be converted to the same frequency as the Stokes photon. This is the first stimulated Raman scattering process. Another photon may arrive at frequency ω_3 that will stimulate the emission of the excitation from the resonance, so that the energy of the photon of frequency ω_3 and the excitation are converted to a new photon of frequency $\omega_4 = \omega_3 + \Delta\omega$, called the "anti-Stokes" photon. The presence of this photon of frequency ω_4 indicates that a CARS process has taken place and indeed a molecule with the resonance feature is present. Often the "pump" beam is used as the photons of frequency ω_3, so that $\omega_3 = \omega_1$ and $\omega_4 = 2\omega_1 - \omega_2$. Since the photon of ω_4 is not the same frequency as one of the illuminating photons, and is typically within the transparency range of the tissue, it is easily discriminated from the incoming radiation. Figure 1 A shows an energy-level diagram for CARS, and Figure 1 B shows an energy-level diagram for Coherent Stokes Raman Scattering (CSRS).

[05] CARS microscopy uses the CARS process to look for the presence of a molecular species, but does not require any foreign substances to be introduced into the tissue. It

scans the illumination point-by-point through the tissue and measures the number of generated anti-Stokes photons. When a three-dimensional mesh of points has been scanned, a complete three-dimensional picture of molecules of that resonance can be shown. Since CARS is a nonlinear process (and therefore is intensity sensitive), efficient conversion only occurs at the focus of the illumination, which can be made very tight (typically a half micron in both the axial and lateral directions). Therefore the resolution can be made many orders of magnitude better than MRI, which is the probably the largest competition for clinical use for similar purposes. Unfortunately, the penetration is usually rather low (limited to about 500 microns). A further shortcoming is that CARS microscopy measures the total number of anti-Stokes photons, or power, from the sample. However, the optical field contains temporal structure in the phase that is averaged out by power detection because photodetector response time is orders of magnitude slower than the oscillations of the optical field. The time scale on which the optical pulse varies (which is typically picoseconds or femtosecond time scales) is far too fast for photon detection equipment or electronics to detect (the fastest of which may detect 25 ps time scales).

[06] Optical coherence tomography (OCT) is an emerging high-resolution medical and biological imaging technology. OCT is analogous to ultrasound B-mode imaging except reflections of low-coherence light are detected rather than sound. OCT detects changes in the backscattered amplitude and phase of light.

[07] Nonlinear interferometric vibrational imaging (NIVI) is a method used to measure the three-dimensional distribution of molecular species in various samples (biological or otherwise) [1]. Its basic operation is to stimulate the excitation of molecular bonds with particular resonance frequencies, and then use these excitations to produce radiation distinct from the excitation that can be measured. Unlike previous methods that use CARS in microscopy to probe for the presence of molecular species, NIVI utilizes a heterodyne approach where a reference signal is separately generated and interferomerically compared to the signal received from the sample, allowing the signal to be formed into an image in the same way as OCT. In this way, additional

information can be inferred from the emitted radiation such as the distance to the sample and phase information that yields additional structure of the molecular bonds. It also has other advantages in sensitivity and the ability to screen out background radiation that is not produced by the sample. It also can allow more flexibility in the choice of laser illumination source, because the coherent detection process does not rely on photon frequency alone to discriminate emitted radiation.

[08] There are compelling reasons to use broadband sources to excite CARS. An ultrafast pulse can be shaped into a longer picosecond pulse that can excite CARS more efficiently. In addition, it can be used to excite many resonances simultaneously [2]. Unfortunately, unlike narrowband pulses, the anti-Stokes radiation produced will not be narrowband. If many resonances are excited simultaneously then the anti-Stokes radiation they produce have overlapping spectra. Because noninterferometric detection can only measure the spectrum of the anti-Stokes radiation, the contributions of each resonance to the anti-Stokes radiation will be inseparable. Interferometric detection allows the demodulation of the anti-Stokes field so the Raman spectrum can be inferred when exciting multiple simultaneous resonances.

[09] In addition, broadband sources should allow for more Raman-frequency agile imaging instruments. When utilizing narrowband pump and Stokes pulses, the frequency difference between them must be tuned to the Raman frequency of interest. Often retuning lasers or amplifiers is difficult to make reliable and automatic. Pulse shaping, however, is achieved by movable gratings, prisms, or mirrors to adjust dispersion and delay, or by acousto-optic or liquid-crystal Fourier plane pulse shapers without moving parts [3]. Because pulse shapers do not typically involve feedback, oscillation, or overly sensitive alignment, pulse shapes can be changed much more easily. In addition, a computer can control these mechanisms automatically, so that changing the pulse shape should be much easier than retuning a laser source.

[10] Two important features of the Raman spectrum in practice that need to be mutually distinguished are the resonant and nonresonant components. The resonant components are specific to features of a molecule, which include vibrational

frequencies, rotational frequencies, and electronic resonances. Nonresonant features are not specific to a particular molecule, and are weakly dependent on frequency.

[11] Distinguishing CARS resonant from nonresonant four-wave-mixing has become problematic with the advent of ultrafast laser sources. These sources produce pulses on the order of 5-200 fs, much shorter than the lifetime of the resonance, which is $1/\Gamma_n$. If transform-limited ultrafast pulses are used, only a small polarization $P^{(3)}(\Omega)$ can be produced in the molecule, while the nonresonant components are enhanced. Since the lifetime of the resonance is typically 1-100 ps, transform-limited ultrafast pulses excite resonant transitions inefficiently and nonresonant transitions efficiently. Before the advent of near-infrared solid-state femtosecond lasers, especially the Ti^{3+}:Al_2O_3 laser, most laser sources produced picosecond-length pulses that favored the generation of CARS. Many current CARS instruments utilize narrowband pump and Stokes pulses of picosecond length for this reason.

SUMMARY

[12] In a first aspect, the present invention is a method of examining a sample, comprising exposing of the sample to a pump pulse of electromagnetic radiation for a first period of time, exposing of the sample to a stimulant pulse of electromagnetic radiation for a second period of time which overlaps in time with at least a portion of the first exposing, to produce a signal pulse of electromagnetic radiation for a third period of time, and interfering the signal pulse with a reference pulse of electromagnetic radiation, to determine which portions of the signal pulse were produced during the exposing of the sample to the stimulant pulse. The first and third periods of time are each greater than the second period of time.

[13] In a second aspect, the present invention is a method of producing an image, comprising collecting a set of data. The data comprises a plurality of digital data produced by the above method.

[14] In a third aspect, the present invention is a method of examining a sample by CARS or CSRS, including exposing a sample to laser light to produce a signal, and producing a data set from the signal. The improvement comprises interfering the signal with a reference pulse, and excluding at least a portion of the signal containing non-resonant electromagnetic radiation in producing the data set.

[15] In a fourth aspect, the present invention is a method of examining a sample, comprising a step for producing a CARS or CSRS signal pulse of electromagnetic radiation from a sample, and a step for determining which portions of the signal pulse contain electromagnetic radiation produced by four-wave mixing.

BRIEF DESCRIPTION OF THE DRAWINGS

[16] Figure 1. Coherent Anti-Stokes Raman Scattering and Coherent Stokes Raman Scattering energy-level diagrams.

[17] Figure 2. Basic block diagram of nonlinear interferometric vibrational imaging (NIVI).

[18] Figure 3. Example laser configurations that produce two pulses of frequencies ω_1 and ω_2 overlapped.

[19] Figure 4. Methods of shaping a broadband pulse into a pulse with beat frequency $\Delta\omega$.

[20] Figure 5. Reference Signal Generator Implementation

[21] Figure 6. Configurations for full field CARS.

[22] Figure 7. Translated serial-point scanning configurations

[23] Figure 8. Beam-steered serial-port scanning configurations

[24] Figure 9. Temporal-ranging based NIVI

[25] Figure 10. Full field cross-correlator demodulator

[26] Figure 11. Temporal cross-correlator for a serial-point scanning microscope.

[27] Figure 12. Two configurations that utilize linear photodetector arrays to measure multiple samples of the cross-correlation simultaneously.

[28] Figure 13. Diagram of an implementation of the recorder.

[29] Figure 14. Electrical cross-correlation signal quadrature amplitude demodulator.

[30] Figure 15. "Biological window" in tissue where absorption of near-infrared wavelengths is at a minimum and light can penetrate deep into highly-scattering tissue.

[31] FIG. 16. Flowchart of preconditioned conjugate gradient algorithm applied to inversion of Raman spectra.

[32] Figure 17. Four-wave-mixing (FWM) and CARS signals produced by a long pump pulse and a short Stokes pulse combination.

[33] Figure 18. Experimentally measured cross-correlations between a short reference pulse and the anti-Stokes radiation of an acetone sample as the pump and Stokes frequency difference are tuned. Away from the resonance at 2925 cm^{-1}, only the nonresonant component exists, while the "tail" is maximized when tuned near resonance.

[34] Figure 19. Schematic of experimental apparatus used to acquire cross-correlation interferograms.

[35] Figure 20. Setup to generate a CARS excitation signal and reference signal to reject nonresonant four-wave-mixing, using a delayed probe pulse so that a reference pulse can be used as an interferometric gate to reject the early-arriving nonresonant four-wave-mixing.

[36] Figure 21. Setup to generate a reference pulse and overlapped dispersed pump and Stokes pulses for measurement of $\chi^{(3)}(\Omega)$.

[37] Figure 22. Setup to measure Raman spectrum in a particular frequency range using an ultrabroadband source.

[38] Figure 23. Graph of instantaneous frequency of two overlapped pulses, where the pulse frequency difference sweeps from Ω_H to Ω_L.

[39] Figure 24. Time/frequency diagram for pulse combination using linearly chirped and delayed pulses.

[40] Figure 25. Simulation of measurement and reconstruction of Raman spectrum. (a) is the magnitude of the illumination pulse as a function of time. (b) is the spectrum of the illumination pulse. (c) is the spectrum of the generated anti-Stokes light. (d) is the spectrum of the beat frequencies of the illumination pulse. (e) is the magnitude of the Raman spectrum of the hypothetical molecule. (f) is the recovered hypothetical spectrum.

[41] Figure 26. Simulation of measurement and reconstruction of Raman spectrum. (a) is the spectrum of the beat frequencies of the illumination pulse. (e) is the magnitude of the Raman spectrum of the hypothetical molecule which includes both resonances. (f) is the recovered hypothetical spectrum.

DETAILED DESCRIPTION

[42] The present invention makes use of the discovery that interferometry may be used to differentiate between resonant CARS and non-resonant four-wave-mixing processes. Inclusion of a reference signal allows for interferometry, and therefore this important source of noise can be eliminated. This may be particularly advantageous when combined with NIVI, which already includes a reference signal.

[43] A possible pump/Stokes combination for measuring spectra using CARS, would include a narrow band pump of frequency ω_1 used with a broadband short Stokes pulse of frequencies $\omega_2 - \Delta\omega/2$ to $\omega_2 + \Delta\omega/2$. The Stokes pulse arrives on the leading edge of the pump pulse, and excites a broad band of Raman frequencies $\omega_2 - \omega_1 - \Delta\omega/2 < \Omega < \omega_2 - \omega_1 + \Delta\omega/2$. Because of the narrowness of the pump, a Raman frequency of Ω will be mapped to an anti-Stokes frequency of $\omega_1 + \Omega$. A spectrometer can then directly sample the anti-Stokes spectrum, which maps onto the Raman spectrum.

[44] While this approach has the appeal of simplicity, it has a number of drawbacks that make it difficult to implement in practice. The chief problem is generating synchronized narrowband pump and broadband Stokes pulses. The bandwidth of the pump must be equal to or less than the bandwidth of the Raman spectral lines probed (on the order of 1-10 cm^{-1}) while the bandwidth of the Stokes must exceed the total bandwidth probed (usually greater than 100 cm^{-1}). Using synchronized picosecond and femtosecond lasers is possible, but difficult and cumbersome. Otherwise, both pulses will usually be derived from a single oscillator, usually a femtosecond oscillator, with an optical parametric amplifier or oscillator used to convert part of the pump pulse to the Stokes pulse. A broadband pump pulse must be used to derive a broadband Stokes pulse. To decrease the bandwidth of the pump for this technique, it can be filtered to a narrow bandwidth. However, because the anti-Stokes output power is proportional to the square of the pump power the decrease in anti-Stokes power will be severe.

[45] If one attempts to start with a narrowband picosecond oscillator, and derive the pump and Stokes from that, the Stokes pulse will need to be broadened using a process such as self-phase-modulation. Unfortunately, these nonlinear broadening processes require very high peak power pulses, which for relatively long picosecond pulses would require very high pulse energy indeed. Also, the bandwidth created in this way tends to be difficult to recompress to a pulse shorter than the original pulse length.

[46] In addition, the Stokes pulse will need to be chirped to avoid overly increasing the nonresonant four-wave-mixing, which will require further lengthening of the pump pulse in time to ensure sufficient Raman frequency resolution. This will further decrease the anti-Stokes output. While sampling of the Raman spectrum is possible using incoherent detection, it presents difficulties that are much more easily solved using coherent interferometric detection.

[47] A typical Raman spectrum $\chi^{(3)}(\Omega)$ can be decomposed into a sum of resonance and nonresonant components:

[48] $$\chi^{(3)}(\Omega)=\chi^{(3)}_{NR}(\Omega)+\sum_n \chi^{(3)}_n(\Omega) \qquad (11)$$

[49] The nonresonant component $\chi^{(3)}_{NR}(\Omega)$ is a slowly changing function of frequency Ω, and is usually approximated by a constant $\chi^{(3)}_{NR}(\Omega)=\chi_{NR}$ which is a real number. The resonant components $\chi^{(3)}_n(\Omega)$ are sharply peaked at the resonance frequency, and are typically described by a homogeneously broadened Lorentzian spectrum:

[50] $$\chi^{(3)}_n(\Omega)=\frac{2\Omega_n}{\Omega^2-2i\Gamma_n\Omega-(\Omega_n^2-\Gamma_n^2)} \qquad (12)$$

[51] where $\Omega=\sqrt{\Omega_n^2-2\Gamma_n^2}$ is the center frequency of the resonance, and Γ_n is the linewidth. A resonance has a Raman magnitude that is sharply peaked around the center frequency. Interferometric detection can distinguish the flat, insensitively dependent on Raman frequency, nonresonant spectra and the sharply peaked resonant CARS spectra. In addition, the imaginary part of $\chi^{(3)}(\Omega)$ is determined by only the resonant Raman component, while the real part is determined by both resonant and nonresonant components. Another of the distinguishing features between a resonant and nonresonant feature is the phase reversal of π that occurs through the center frequency. Noninterferometric instruments cannot distinguish the phase, while the interferometric method measures the complex $\chi^{(3)}_n(\Omega)$. As well as being a signature of

CARS, this phase reversal may be used to separate several resonances close together in frequency.

[52] To see how interferometry can detect the difference between resonant and nonresonant processes, consider the shape of the anti-Stokes pulses produced by the long pump and short Stokes pulses previously mentioned. This combination is illustrated in Figure 17. When the pump and Stokes pulses overlap, the molecule will be excited by stimulated Raman scattering (SRS). This excitation will remain after the Stokes pulse passes. At the moment of the overlap, nonresonant four-wave-mixing processes can also be excited. However, because there is no persistent state associated with nonresonant processes, the emission of four-wave-mixing will end quickly after the Stokes pulse passes. With a Raman-active resonance, however, the pump can produce anti-Stokes radiation via SRS even after the Stokes has passed, because the resonance persists. Therefore, the nonresonant component can be discarded by rejecting any anti-Stokes radiation that occurs coincident with the Stokes pulse.

[53] The benefit of interferometry is that the time of arrival of the anti-Stokes radiation can be found very precisely by cross-correlating a broad bandwidth reference pulse with the anti-Stokes radiation. Incoherent detection can detect the interference between the resonant and nonresonant components of the anti-Stokes radiation, but does not directly detect the time of arrival of the anti-Stokes light. Figure 18 shows the actual measured cross-correlations between a short reference pulse and the anti-Stokes radiation for a cuvette of acetone with a resonance at 2925 cm^{-1} as the frequency difference between the pump and Stokes is tuned.

[54] The apparatus to acquire the interferogram is documented in Figure 19. A regenerative amplifier produces a pulse at 808 nm and 30 nm bandwidth that is used as a seed pulse for a second-harmonic-generation optical parametric amplifier, and also as the pump for the CARS sample. The pump is lengthened to approximately 200 fs by passing it through a dispersive Dove prism made of BK7 glass of 105 mm length. The idler of the optical parametric amplifier produces a Stokes pulse at 1056 nm and 70 fs in length, which is combined with the pump at a dichroic mirror. The pump is delayed

so the Stokes arrives on the leading edge of the pump pulse. The signal of the optical parametric amplifier is at 653 nm and serves as the reference pulse because it matches the frequency of the anti-Stokes produced in the CARS sample when the pump and Stokes are focused into it. The anti-Stokes and reference are cross-correlated by delaying them relative to each other and detecting their interference power on a silicon photodetector.

[55] Note that to detect the difference between resonant and nonresonant processes did not require that the pump be lengthened to the lifetime of the resonance of the acetone. This is because the nonresonant component ends abruptly at the end of the Stokes pulse. Unlike the incoherent detection case, one can actually reject the nonresonant component based on its time of arrival, and not just depend on minimizing the amount of nonresonant four-wave-mixing that is generated.

[56] Instead of dispersing the pump pulse to create a portion of the pump that can act as a probe to produce SRS in the sample, the pump pulse can be split into two and one part delayed to act as the probe. This can be achieved with the setup in Figure 20. A regenerative amplifier or a mode-locked laser oscillator generates short pulses which are split by a beamsplitter into two pulses. One pulse is to create the pump/probe pulses. This pulse is further split in two, and one copy is delayed with respect to the other. The earlier pulse is the pump and the later pulse the probe. The other pulse not used to create the pump/probe pulses is used in an optical parametric amplifier or an optical parametric oscillator to generate nearly transform-limited idler and signal pulses. The idler is used as the Stokes pulse, and the signal is used as the reference pulse. The idler is delayed so that it overlapped with the pump pulse in time and space with a dichroic beamsplitter. The pump/Stokes/probe combination is sent to a microscope to excite the sample. The probe will create stimulated Raman scattering, which is collected by the microscope. The signal is used as the reference and is delayed to arrive at the interferometric demodulator at the same time as the SRS created by the probe. The magnitude of the interference signal indicates the presence of the target species in the sample.

[57] In some cases, the optical parametric amplifier or oscillator may not provide a signal at a suitable frequency to use as a reference pulse. If a second-harmonic-generation optical parametric amplifier (SHG OPA) is used, then the sum of the idler and signal frequencies will equal two times the seed frequency, and so the signal will be of the correct frequency. If a standard optical parametric amplifier or oscillator is used, then the signal and idler frequencies add up to the seed frequency. The signal pulse can be used as the Stokes pulse, but the idler frequency is too low to be useful. In this case, one can produce a reference by combining nearly transform-limited pump and Stokes pulses with four-wave-mixing. Effectively a SHG OPA/OPO implements a four-wave-mixing process with two three-wave-mixing processes, but the same can be done directly in a nonresonant four-wave-mixing medium such as quartz, sapphire, CaF_2, LiF, water, or any medium with a sufficiently strong four-wave-mixing cross section, transparency to the pump/Stokes/anti-Stokes wavelengths, and no nearby resonant Raman frequencies.

[58] We now present a method using a pulse sequence like that of Figure 17 to recover the Raman spectrum $\chi^{(3)}(\Omega)$ in a particular frequency range. The method uses a pump that is broadband but dispersed to be lengthened in time, while using a short, nearly transform-limited Stokes pulse. These two pulses will be overlapped so that the Stokes pulse is on the leading edge of the pump pulse. Figure 21 shows a block diagram of apparatus that can produce this pulse combination:

[59] The idea of the setup of Figure 21 is to stimulate a broad bandwidth of Raman transitions using a short Stokes pulse overlapped with the leading edge of a pump pulse. After the Stokes pulse is over, the pump pulse continues to produce stimulated Raman scattering, which is demodulated with the reference pulse. The portion of the cross-correlation that overlaps the Stokes pulse arrival can be discarded to remove the nonresonant four-wave-mixing component of the anti-Stokes signal. Because the field used to produce the resonant cars is only due to the pump pulse, the cross-correlation amplitude and the pump pulse can be used to estimate the Raman spectrum $\chi^{(3)}(\Omega)$ (for example, using an inverse of Eq. 6 in the Example section, below).

[60] To elaborate on the setup, a regenerative amplifier or oscillator generates a train of ultrashort pulses, typically several hundred cm^{-1} in bandwidth. These pulses are split into two by a beam splitter. One of the pulses is used in a dispersive pulse shaper, which imparts a non-constant group delay to the pulse. Such a dispersive pulse shaper can be implemented by dispersive materials such as optical glasses or other transparent materials, prism or grating type pulse shapers, or optical fibers or waveguides. The pulse will usually be dispersed to a length on the order of the lifetime of the resonances under study. The other pulse will be sent to an optical parametric amplifier or oscillator, where it will generate a broadband signal and idler pulse. The idler will act as the Stokes and will be combined with a dichroic beamsplitter with the dispersed pulse at its leading edge. This combination will illuminate the sample through the microscope, and will have anti-Stokes radiation collected by the microscope. The signal will act as a reference pulse and will be cross-correlated with the anti-Stokes light generated by the sample. For best performance, the pump/Stokes signal should be filtered out of the anti-Stokes signal before cross-correlation.

[61] In addition, there is an optional section of Figure 21 that helps enhance the amount of excitation generated in the sample. Instead of sending all of the pump light into the dispersive pulse shaper, it can be further split in two, where a portion will enter the pulse shaper and become a probe pulse. The other portion will be a pump pulse, and is delayed but not dispersed and is recombined with the dispersed pulse. The undispersed pump pulse will be delayed to overlap the Stokes pulse on the leading edge of the dispersed probe pulse. By increasing the intensity of the pump at the instant the Stokes arrives, the magnitude of the excitation of the resonances can be enhanced, allowing a larger stimulated Raman signal to be collected from the sample.

[62] A theory of CARS interferometry and inference, including Eq. 1-10, is presented in the EXAMPLES section, below.

[63] To understand in better detail how the present invention can recover the Raman spectrum, consider a dispersed pump pulse and a Stokes pulse with an electric field

[64] $$\tilde{E}_p(\omega) = A_p(\omega)\exp(i\phi(\omega)) \text{ for } \omega_0 < \omega < \omega_0 + \Delta\omega_P \quad (13)$$

[65] $$\tilde{E}_p(\omega) = 0 \text{ otherwise}$$

[66] $$\tilde{E}_s(\omega) = A_s \text{ for } \omega_0 - \Omega_0 - \Delta\omega_S/2 < \omega < \omega_0 - \Omega_0 + \Delta\omega_S/2 \quad (14)$$

[67] $$\tilde{E}_s(\omega) = 0 \text{ otherwise}$$

[68] The frequency Ω_0 corresponds to the resonant Raman frequency of interest, while the bandwidths $\Delta\omega_P$ and $\Delta\omega_S$ give the bandwidths of the pump and Stokes signals respectively. The amplitudes $A_p(\omega)$ and A_s correspond to the real, positive amplitudes of the frequency components of the pump and Stokes signals, respectively. To find these pulses in the time domain, we use the stationary phase approximation (which should apply well to dispersed pulses). We define $d\phi/d\omega = t'(\omega)$, $\omega'(t) = t'^{-1}(\omega)$ (the inverse function of $t'(\omega)$), and $d\Phi/dt = \omega'(t)$. Note that $t'(\omega)$ must be a strictly increasing or decreasing function of ω. The time-domain versions of these signals are:

[69] $$E_p(t) = A_p(\omega'(t))\left|\frac{d^2\Phi}{dt^2}\right|^{-1/2}\exp(i\Phi(t)) \text{ for } t'(\omega_0) < t < t'(\omega_0 + \Delta\omega_P) \quad (15)$$

[70] $$E_t(\omega) = 0 \text{ otherwise}$$

[71] $$E_s(t) = A_s\exp(i(\omega_0 - \Omega_0)t)\sin(t\Delta\omega_S)/(t\Delta\omega_S) \quad (16)$$

[72] To place the Stokes pulse on the leading edge of the pump pulse, we assume that $d\phi/d\omega|_{\omega=\omega_0} = 0$. At time t=0, the Raman frequencies $\Omega_0 - \Delta\omega_S/2 < \Omega < \Omega_0 + \Delta\omega_S/2$ are excited. Because the Stokes is nearly transform-limited, the phases of Eq. 1 cancel and the polarization is proportional to the Raman spectrum:

[73] $$P^{(3)}(\Omega) \propto E_p E_s \chi^{(3)}(\Omega) \text{ for } \Omega_0 - \Delta\omega_S/2 < \Omega < \Omega_0 + \Delta\omega_S/2 \quad (17)$$

[74] $$P^{(3)}(\Omega) = 0 \text{ otherwise}$$

[75] The anti-Stokes generated from this polarization is:

[76] $$\tilde{E}_A(\omega) = \int_{\Omega_0-\Delta\omega_S/2}^{\Omega_0+\Delta\omega_S/2} P^{(3)}(\Omega) A_p(\omega-\Omega)\exp(i\phi(\omega-\Omega))d\Omega \qquad (18)$$

[77] If we define $P^{(3)}(t) = \int_{\Omega_0-\Delta\omega/2}^{\Omega_0-\Delta\omega/2} P^{(3)}(\Omega)\exp(-i\Omega t)dt$ then this becomes a product:

[78] $$E_A(t) = P^{(3)}(t)A_p(\omega'(t))\left|\frac{d^2\Phi}{dt^2}\right|^{-1/2}\exp(i\Phi(t)) \quad \text{for } t'(\omega_0) < t < t'(\omega_0+\Delta\omega_P) \qquad (19)$$

[79] where $E_A(t)$ is the complex analytic continuation of the Fourier transform of $\tilde{E}_A(\omega)$. This suggests that $P^{(3)}(t)$ can be recovered by multiplying $E_A(t)$ by the conjugate of the phase of the probe field $\exp(-i\Phi(t))\left|\frac{d^2\Phi}{dt^2}\right|^{1/2}$ in the time domain. The $P^{(3)}(\Omega)$ and therefore $\chi^{(3)}(\Omega)$ can be recovered from $P^{(3)}(t)$ by means of a Fourier transform.

[80] Raman spectra retrieval using broadband pulses

[81] One of the chief benefits of the invention is that enables the use of ultrabroadband lasers with thousands of cm^{-1} of bandwidth to be used to measure either narrow (less than 10 cm^{-1}) or a very wide range of Raman frequencies. Such lasers include ultrabroadband Ti-sapphire lasers, Cr:forsterite lasers, dye lasers, Yb:YAG, Yb-silica fiber, Er-silica fiber, Nd:glass, and femtosecond lasers (Ti-sapphire, dye lasers, Cr:forsterite, Yb:YAG, Yb- silica fiber, Er- silica fiber, Nd:glass) broadened by continuum generation (e.g. using bulk materials, such as photonic crystal fibers, high numerical aperture fibers, or microstructured optical fibers, as well as sapphire and water). The benefits of this are that the same laser can be used to produce light that both acts as the pump and Stokes frequencies (rather than requiring a separate Stokes to be derived), and the entire power spectral bandwidth of the laser can be used to contribute to the excitation and stimulated Raman scattering of the sample. When using

narrowband pulses, the power spectral density of the pump and Stokes must be very high to ensure sufficient excitation. With ultrabroadband excitation one can use its relatively low power spectral density to excite, but it does not produce a one-to-one correspondence between anti-Stokes and Raman frequencies, so that interferometric detection is needed.

[82] The essence behind these methods is a reinterpretation of Eq. 1. In the time domain, Eq. 1 becomes:

[83] $$P^{(3)}(t) = \int_0^\infty \chi^{(3)}(\tau)\left|E_i(t-\tau)\right|^2 d\tau \qquad (20)$$

[84] Eq. 20 is the causal convolution of the impulse Raman response $\chi^{(3)}(t) = \int_{-\infty}^{\infty} \chi^{(3)}(\Omega)\exp(-i\Omega t)dt$ with the instantaneous intensity of the incoming pulse. If there is a modulation of the intensity at the same frequency as a Raman resonance, the Raman resonance will become excited. By using a pulse that contains a modulation of the intensity of the excitation wave that spans a range of frequencies, an entire range of the same Raman frequencies can be excited and studied. Previously, narrowband lasers were the primary source available for stimulating CARS, so that the intensity modulation was created by the beats between two narrowband pulses. Broadband sources have the benefit of providing a much more easily tunable source of beats of a particular frequency, because the laser need not be retuned, only the pulse shaped differently.

[85] An ultrabroadband source of pulsed radiation can be used to both stimulate CARS and act as the reference. The pulses from the source will be divided by a frequency-selective element such as a dichroic beam splitter into lower and higher frequency pulses. The higher frequency pulse bandwidth will correspond to the anti-Stokes frequencies emitted by the sample, and will act as the reference pulse to demodulate the anti-Stokes signal. The lower frequency pulse will be shaped to

stimulate the Raman frequencies of the sample in a particular bandwidth. This scheme is illustrated in Figure 22.

[86] The strategy used here is to split the lower frequency pulse into two copies. Each copy will be put through separate dispersive elements, and then recombined afterwards with a time delay between them. Dispersive elements can be implemented by one or a combination of dispersive transparent materials (e.g. optical glasses or liquids), prism or grating pulse compressors or expanders, spatial-light-modulator based pulse shapers (e.g. liquid-crystal or acousto-optic Fourier plane pulse shapers), or dispersive optical fibers or waveguides. Dispersion causes the various frequency components of the pulses to be spread out in time. When the two pulses are overlapped, a beat frequency is produced at the difference between the instantaneous frequencies of the two pulses at a given instant. By causing the frequency difference of the two pulses to vary between a lower and higher difference frequency during the time interval they are overlapped, the beats can stimulate the Raman frequencies in the same range. By varying the instantaneous frequency of the two pulses, and their relative delay, the range of Raman frequencies the overlapped pulses stimulate can be varied.

[87] To see how two dispersed pulses can produce a beat frequency spectrum that ranges from $\Omega_L < \Omega < \Omega_H$, consider two band limited pulses with different dispersion phases imposed on them:

[88] $$\tilde{E}_1(\omega) = E_1 \exp(i\phi_1(\omega)) \text{ for } \omega_0 - \Delta\omega/2 < \omega < \omega_0 + \Delta\omega/2 \qquad (21)$$

[89] $$\tilde{E}_1(\omega) = 0 \text{ otherwise}$$

[90] $$\tilde{E}_2(\omega) = E_2 \exp(i\phi_2(\omega)) \text{ for } \omega_0 - \Delta\omega/2 < \omega < \omega_0 + \Delta\omega/2 \qquad (22)$$

[91] $$\tilde{E}_2(\omega) = 0 \text{ otherwise}$$

[92] The dispersion phase $\phi_1(\omega)$ and $\phi_2(\omega)$ correspond to the total phase a frequency ω of its respective pulse accumulates in its respective dispersive pulse shaper. For example, if pulse #1 travels through a medium with dispersion relation $k(\omega)$ and

thickness d, then the dispersion phase for that pulse $\phi_1(\omega) = k(\omega)d$. In the stationary phase approximation, the time-domain signals for these pulses are:

[93] $$E_1(t) = E_1\left|\frac{d^2\Phi_1}{dt^2}\right|^{-1/2} \exp(i\Phi_1(t)) \text{ for } t_1'(\omega_0 - \Delta\omega/2) < t < t_1'(\omega_0 + \Delta\omega/2) \quad (23)$$

[94] $$E_1(\omega) = 0 \text{ otherwise}$$

[95] where $\frac{d\phi_1}{d\omega} = t_1'(\omega)$, $\omega_1'(t) = t_1'^{-1}(\omega)$, and $\frac{d\Phi_1}{dt} = \omega_1'(t)$.

[96] $$E_2(t) = E_2\left|\frac{d^2\Phi_2}{dt^2}\right|^{-1/2} \exp(i\Phi_2(t)) \text{ for } t_2'(\omega_0 - \Delta\omega/2) < t < t_2'(\omega_0 + \Delta\omega/2) \quad (24)$$

[97] $$E_t(\omega) = 0 \text{ otherwise}$$

[98] where $\frac{d\phi_2}{d\omega} = t_2'(\omega)$, $\omega_2'(t) = t_2'^{-1}(\omega)$, and $\frac{d\Phi_2}{dt} = \omega_2'(t)$.

[99] We wish to further restrict the time interval of the overlap be confined from $-T/2 < t < T/2$. To do this and utilize the full bandwidth of the signal we force pulse 1 to end at time $T/2$, and force pulse 2 to begin at time $-T/2$:

[100] $$\left.\frac{d\phi_1}{d\omega}\right|_{\omega=\omega_0+\Delta\omega/2} = \frac{T}{2} \text{ and } \left.\frac{d\phi_2}{d\omega}\right|_{\omega=\omega_0-\Delta\omega/2} = \frac{-T}{2} \quad (25)$$

[101] We would also like the frequency difference between the two pulses to start at Ω_H and end at Ω_L. This best places the anti-Stokes frequencies outside of the bandwidth of the pump and Stokes frequencies:

[102] $$\left.\frac{d\Phi_1}{dt}\right|_{t=-T/2} - \left.\frac{d\Phi_2}{dt}\right|_{t=-T/2} = \Omega_H \text{ and } \left.\frac{d\Phi_1}{dt}\right|_{t=T/2} - \left.\frac{d\Phi_2}{dt}\right|_{t=T/2} = \Omega_L \quad (26)$$

[103] $$\Omega_L < \left[\frac{d\Phi_1}{dt} - \frac{d\Phi_2}{dt}\right] < \Omega_H \text{ for } \frac{-T}{2} < t < \frac{T}{2} \quad (27)$$

[104] $$\frac{d^2\Phi_2}{dt^2} > \frac{d^2\Phi_1}{dt^2} > 0 \quad (28)$$

[105] The inequalities of Eq. 27 ensure that the beat frequency remains inside the excitation interval. The instantaneous frequency of two pulses that fit these criteria are shown in Figure 23:

[106] Based on the pulses of Eqs. 21 and 22, we can determine a method of finding the CARS signal. Based on Eqs. 1 and 2, these will be:

[107] $$P^{(3)}(\Omega)=\chi^{(3)}(\Omega)\int_{\omega_0-\Delta\omega/2}^{\omega_0+\Delta\omega/2-\Omega}E_1E_2^{*}\exp(i\phi_1(\omega+\Omega)-i\phi_2(\omega))d\omega \qquad (29)$$

[108] $$\tilde{E}_A(\omega)=\int_0^{\omega}P^{(3)}(\Omega)[E_1\exp(i\phi_1(\omega-\Omega))+E_2\exp(i\phi_2(\omega-\Omega))]d\Omega \qquad (30)$$

[109] The last equation can be recast as:

[110] $$E_A(t)=P^{(3)}(t)\left[E_1\left|\frac{d^2\Phi_1}{dt^2}\right|^{-1/2}\exp(i\Phi_1(t))+E_2\left|\frac{d^2\Phi_2}{dt^2}\right|^{-1/2}\exp(i\Phi_2(t))\right] \qquad (31)$$

[111] The inversion in all cases including that of Eqs. 29 and 30 can be implemented by the regularized least-squares solution already described. In general, because the anti-Stokes radiation generated by pulse #1 and pulse #2 overlap in frequency, a solution like that implemented for Eq. 15 will not suffice for this case.

[112] We note that if the lowest Raman frequency Ω_L is stimulated exactly at the end of the overlap of the two pulses at time $t=T/2$, then there is no time to accumulate anti-Stokes signal from that Raman frequency. Therefore in practice one should choose an Ω_L slightly lower than the minimum desired Raman frequency, perhaps decreased by 20% of $\Omega_H-\Omega_L$. This ensures that the upper frequency pulse continues long enough to read out the minimum desired Raman frequency.

[113] It is also possible to reverse the beat frequency so that the frequency is swept from Ω_L to Ω_H. To do this, one can pose the following constraint on the signals:

[114] $$\left.\frac{d\Phi_1}{dt}\right|_{t=-T/2} - \left.\frac{d\Phi_2}{dt}\right|_{t=-T/2} = \Omega_L \text{ and } \left.\frac{d\Phi_1}{dt}\right|_{t=T/2} - \left.\frac{d\Phi_2}{dt}\right|_{t=T/2} = \Omega_H \qquad (32)$$

[115] $$\frac{d^2\Phi_1}{dt} > \frac{d^2\Phi_2}{dt} > 0$$

[116] with the added constraint of Eq. 27. In addition, one can filter the output signal from the sample for frequencies less than the minimum frequency of the illumination $\omega_0 - \Delta\omega/2$. In this case, one captures the CSRS output of the sample. A reference pulse of this same frequency band will then demodulate this signal, and a regularized least-squares inversion operator for Eq. 7 can be employed to find the Raman spectrum.

[117] As an example of a pulse combination that can stimulate a band of Raman frequencies [2], we design a pulse that sweeps the beat frequency linearly from Ω_L to Ω_H. To do this, we will split a transform-limited pulse of bandwidth $\omega_0 - \Delta\omega/2 < \omega < \omega_0 + \Delta\omega/2$ into two components. Each component will have a linear chirp (quadratic phase) applied to it in the frequency domain. This chirp is imparted by the dispersive pulse shaper elements of the device. To impart a linear chirp, a pulse dispersing system that could consist of dispersive materials (e.g. optical glasses or crystals), prism compressors and expanders, grating compressors and expanders, dispersive waveguides and optical fibers, and Fourier-plane pulse shapers will be needed. Producing a linear chirp will likely require cascaded positive and negatively dispersive elements to cancel the higher dispersion orders (e.g. cubic and quadric dispersion). The design and characterization of particular phase profiles $\phi(\omega)$ is well documented in the literature.

[118] The spectrum of the pulse combination that produces beats from Ω_L to Ω_H over an interval T is given by:

[119] $$\tilde{E}_i(\omega) = E_0 \cos\left(\frac{\pi(\omega-\omega_0)}{\Delta\omega}\right)\left[\left(\frac{1+\kappa}{2}\right)\exp\left(\frac{-i(\omega-\omega_0)\tau}{2} - \frac{i(\omega-\omega_0)^2}{2(\alpha-\beta)}\right)\right. \qquad (33)$$

[120] $$+\left(\frac{1-\kappa}{2}\right)\exp\left(\frac{i(\omega-\omega_0)\tau}{2}-\frac{i(\omega-\omega_0)^2}{2(\alpha+\beta)}\right)\Bigg] \text{ for } \omega_0-\frac{\Delta\omega}{2}<\omega<\omega_0+\frac{\Delta\omega}{2}$$

[121] $\tilde{E}_I(\omega)=0$ otherwise

[122] where $\alpha=(2\Delta\omega-\Omega_H-\Omega_L)/2T$, $\beta=(\Omega_H-\Omega_L)/2T$, and $\tau=(T/2)[\Omega_H/(\Delta\omega-\Omega_H)+\Omega_L/(\Delta\omega-\Omega_L)]$. A cosine apodization window has been added to improve the stability to the inverse. The constant α is the common chirp to both pulses, β is the difference chirp between both pulses, τ is the time delay between the pulses, and κ is the difference in field magnitude between the pulses. To be able to form beats at all Raman frequencies, $\Delta\omega>\Omega_H$. Figure 24 shows a graph of the frequency vs. time for this pulse.

[123] As illustrated in Figure 2, a preferred embodiment NIVI **200** has the following components [1]:

[124] **Oscillator.** This oscillator **201** produces an optical field that can excite the resonance mode of the target molecule through a nonlinear technique (usually stimulated Raman scattering), and also the photon to stimulate the photon that is measured (also usually through stimulated Raman scattering). The combination of these processes is called CARS.

[125] **Reference Generator.** The reference signal generator **203**, which can sometimes be incorporated into the oscillator, converts part of the oscillator signal to a reference signal that can be used in the demodulator section. It acts as a known signal that demodulates the unknown signal from the sample in the interferometer.

[126] **Microscope.** The microscope **205** delivers the field produced by the oscillator to the sample, and collects the field emitted by the sample. The excitation field is usually delivered by a microscope objective, where the oscillator field is focused tightly or sparsely, depending on the scanning method. This focus is scanned through the tissue, and based on the signal received from each tissue volume an image can be formed. When the oscillator signal is delivered to the tissue containing a molecule with a

compatible resonance, a nonlinear process such as CARS can occur and produce a new sample signal (called the "anti-Stokes" for CARS processes). This sample signal serves as an indicator of the presence of the molecular resonance, and also provides additional information about the molecule through the temporal structure of the sample signal.

[127] **Demodulator.** The demodulator **207** combines the signal received from the sample with the reference signal. This is typically achieved by constructing an interferometric cross-correlator. The cross-correlation of the two signals is then measured by a single photodetector or array of photodetectors. The power received by these photodetectors allows the cross-correlation signal to be inferred, from which the temporal response signal from the sample can be also inferred. With knowledge of the physics of the molecule, the presence of and potentially properties of the molecule being tested can be inferred from its temporal response.

[128] **Recorder.** The data recorder **209** records the data measured by the demodulator. This data can be digitally processed to produce an image that a human operator can interpret.

[129] Each of these modules can be implemented in a variety of different ways that can be tailored to various data acquisition needs. In addition, while these units are the basic units of the invention, often the parts can be consolidated to simplify implementation or reduce cost. While the basic block design could be construed as that of a standard interferometric microscope, nonlinear processes are occurring in the reference generator and "Microscope" sections that allow the resonance information of the sample to be pumped.

[130] Each of these units will be detailed presently.

[131] 1. **Oscillator.**

[132] The oscillator produces the electromagnetic field that stimulates the excitation of the resonance to the probed. It also provides the photon that stimulates the output photon that is detected as evidence of CARS or CSRS. There are many types of

oscillators and fields that can produce CARS. Each pulse produced by the oscillator should be nearly identical so that it can excite consistent signals in the reference generator and sample. If the oscillator produces too variable of a signal, the signals from the reference generator and sample may change and produce signals that can be confused with noise sources. Variability in the oscillator output is a noise source in itself that adds uncertainty to what the expected demodulated signal should be.

[133] The conventional way to produce CARS is to send in two overlapped optical pulses, one of which at frequency ω_1, the pump, and the other at ω_2, the Stokes pulse, where $\omega_1-\omega_2= \Delta\omega$, where $\Delta\omega$ is the resonance frequency of the molecules of interest. These pulses produce a beat frequency of $\Delta\omega$ that manifests itself in the magnitude of the optical field. In linear time-independent optics, systems are sensitive only to the frequencies of the optical pulses themselves, and not any beats they may produce together. However, with sufficient intensity the intensity envelope may itself stimulate the molecule. By choosing two pulses that produce beats of this frequency, we can stimulate the molecule with two wavelengths that the tissue is transparent to. Once the resonance is stimulated, another photon of frequency ω_1 (in CARS), or of frequency ω_2 (in CSRS) stimulates the emission of a fourth photon, which is of frequency $2\omega_1-\omega_2$ for CARS, and $2\omega_2-\omega_1$ for CSRS.

[134] Systems that can be used to produce these two frequencies are shown below in Figure 3. A common configuration to produce pulses of these two wavelengths that are overlapped in time is to have a pulsed laser produce one of these wavelengths, split off of some of its energy, and use this energy to produce a second pulse of a lower or higher frequency. In one configuration **300**, a pump laser **301** pumps a dye laser **303**, for example a doubled Nd:YAG pump laser at 532 nm pumping a tunable dye laser. In another possible configuration **302**, a pump laser such as a Ti-sapphire oscillator pumps an optical parametric oscillator (a device that converts pulses to lower frequencies) **305**.

[135] In yet another configuration **304**, the pump laser pumps a regenerative amplifier **307**, such as a Ti-sapphire regenerative amplifier. The regenerative amplifier then pumps an optical parametric amplifier (another frequency conversion device) **309**.

Alternatively, as illustrated in configuration **306**, the pulses of each wavelength are generated by two separate pump lasers, and the time overlap is maintained by a circuit **311** that synchronizes the two sources. In another configuration **308**, the pump laser pumps a continuum light generator **313**, generating broadband light which is filtered by a filter for the two wavelengths with group velocity dispersion correction **315**.

[136] While directly generating the two frequencies and superimposing them to produce beats is the most common way to stimulate CARS, this method has some disadvantages for the method of NIVI. In CARS and CSRS, there are two types of generated signals. Resonant signals depend on the presence of a molecule of a particular resonance frequency to be present to generate the CARS signal. Another component, nonresonant CARS, does not require a particular frequency to perform conversion. Nonresonant CARS depends on the peak intensity in the signal, while the resonant component can build up from many beat periods and so therefore can be spread out in time. Because of this, it is advantageous to spread the CARS signals in time to reduce the nonresonant component.

[137] However, when the two signals are discretely generated and are transform-limited (are not chirped in time), the only way to broaden the signals in time is to reduce their bandwidth. To achieve sufficient power-spectral-density to cause efficient conversion, the pulses must either generated by a low-bandwidth laser, or much power will be wasted in filtering a higher bandwidth signal. Unfortunately, the range resolution in OCT, when temporal ranging is used, is inversely proportional to the illumination bandwidth. This requirement for high bandwidth conflicts with the requirement for small bandwidth for resonance specificity. It would be desirable to come up with an alternate oscillator configuration that would preserve the resonance specificity of the low bandwidth pulses, but actually utilize high bandwidth signals.

[138] Since the nonlinear excitation of the resonant molecule depends more on the beats produced than on the bandwidth used to produce them, it would be desirable to take a broadband pulse and reshape it into a signal with the required beat frequency. Recent advancements have made pulsed sources of very large bandwidth. Some of the

methods to do this are high-bandwidth Ti-sapphire oscillators, dispersion compensated mirror Ti-sapphire oscillators, double chirped-mirror Ti-sapphire oscillators, and continuum generation sources. The optical field produced by these sources can be shaped into a field with the beats at the required frequency.

[139] One such method that has been demonstrated in the literature is shown in Figure 4. A source of laser pulses from a laser source **401** is sent into a Fourier-plane pulse shaper **403** that utilizes a spatial-light-modulator (e.g. liquid crystal modulator or acousto-optic modulator). The Fourier-plane pulse shaper enables each frequency in the pulse to have its phase and/or amplitude altered. By applying the correct phase and amplitude to each incoming frequency, the incoming signal can be convolved with an essentially arbitrary signal. The pulse shaper is set up to reshape the incoming pulse by applying a period phase or amplitude perturbation in the Fourier domain with period $\Delta\omega$/N, where N is a positive integer. This will transform a single pulse into a train of pulses that are separated in time by $2\pi/\Delta\omega$. If only a phase perturbation is used, the power of the signal can be maximally preserved. The larger the integer N is, the longer the pulse train will be, and none of the bandwidth of the original pulse will be lost. However, most pulse shapers have a limited number of controllable frequencies, limiting the practical size of N.

[140] One advantage of spatial-light-modulator based pulse shapers is that there is typically a wide range of pulse shapes that can be achieved, and the spatial-light-modulator can often be controlled automatically by a computer. The computer can then adjust the spatial-light-modulator to achieve maximum signal from the sample in a feedback loop. This may allow automatic correction of dispersion or aberrations introduced by the optics of the system, and will permit more flexibility in probing the molecule because of the tunability of the pulse shapes.

[141] The pulse-shaper in the "Pulse-shaper type NIVI oscillator" **400** is well described in the literature. It consists of two diffraction gratings, which disperse and recombine the frequencies, two lenses that focus each frequency to a point and recollimate each frequency, and a pulse shaper placed at the focal plane to separately operate on each

frequency. The pulse shape is altered by dispersing each frequency to a separate angle, and then imaging each frequency to a separate point on the spatial-light-modulator. Alternatively, an etalon may be used to shape the amplitude of the pulse periodically. Unfortunately, while this would be simpler, it modifies the spectrum of the pulse and therefore introduces artifacts into the NIVI image.

[142] An alternative method is to take a pulse and impart a linear chirp to it. A linear chirp turns a pulse into one where the frequency rises or falls at a linear rate as a function of time. This rate is characterized by a constant α, which is the change in frequency per unit time. It is called "chirped" because of the noise of the equivalent sound wave. If two copies of the chirped pulse are created, delayed with respect to each other by imparting a variable time delay **407**, and recombined, the resulting pulse will have two simultaneous frequencies that will rise or fall together at the same linear rate, but always be separated at a given instant by the same frequency. If this separation frequency is chosen to be $\Delta\omega$, then the envelope of the pulse will be modulated by beats of this frequency. This method is especially convenient because the probed resonance frequency can be adjusted easily by adjusting the time delay between the two chirped pulses, which is relatively easy and inexpensive. This will enable a NIVI instrument that can be rapidly and easily adjusted to scan a wide range of molecular resonances. Systems based on tunable frequency sources will likely be much more difficult to dynamically change reliably and often.

[143] The ability to linearly chirp a pulse is well known in the literature. It can be accomplished with a pulse shaper **405** having a combination of prisms, diffraction gratings, lenses, mirrors, and/or dispersive materials. Combinations may be required to ensure that the resulting chirp is linear and does not contain significant amounts of higher-order dispersion. Higher-order dispersion would limit the resolution to which the resonance could be addressed and exclude other nearby frequency resonances. In a typical Chirped CARS NIVI setup **402**, the chirp rate required would be fixed and the chirp rate should need little or no adjustment in the field. Measuring devices such as Frequency Resolved Optical Gating can test whether a chirped pulse is linearly chirped.

[144] When using high-bandwidth excitation for CARS, it is important to filter out the entire bandwidth of excitation before detection so it does not interfere with detection of the emitted anti-Stokes light (Stokes for CSRS), because the nonlinear emission can not be easily distinguished from the much larger linearly scattered excitation light. However, this linearly scattered light contains the same structure that conventional OCT imaging does, and may be used to measure this information at the same time that a NIVI image is recorded. This will be convenient for superimposing OCT and NIVI data onto the same image, because the acquisition of both types of data can be designed into the same instrument. This can be implemented in practice by using a dichroic beamsplitter to separate the excitation and response radiation, and measuring the cross-correlation of the two frequency bands separately using a cross-correlation demodulator.

[145] With high-bandwidth sources where the entire bandwidth need not be utilized to produce the excitation field, it is possible to use the upper end (for anti-Stokes) or lower end (for Stokes) of this bandwidth as a reference field, eliminating the need to separately generate a reference field. However, the frequencies of the reference must occupy the same band as the received CARS/CSRS light from the sample. Some sources, especially continuum generation sources, will likely generate much more bandwidth than needed for pulse shaping and therefore will probably have this extra bandwidth available for this use. While a different process from CARS/CSRS typically generates this light, it will likely remain phase-coherent with the CARS/CSRS light and therefore should be useful as a reference. Phase-coherence depends on the mechanism of pulse/continuum generation and therefore its phase-coherence stability properties of a particular source type must be established before it is suitable for this purpose. A dichroic beamsplitter may be used to separate the frequency band corresponding to the response radiation from the oscillator energy, so that it may be utilized as a reference signal.

[146] It is also possible to simultaneously stimulate the excitation of several resonances if the sample is illuminated with the appropriate pump and Stokes/anti-Stokes beams. For example, in CARS a narrowband pump signal and a wideband Stokes can be used

to address many resonances simultaneously. This is called multiplex CARS and can be extended to CSRS with a broad anti-Stokes wavelength range. This may be used to measure the presence of several molecular resonances simultaneously in the sample. In addition, if several excitations can be produced in the same molecule simultaneously, the molecule will evolve to various quantum states depending on the relative amplitude and timing between the CARS/CSRS stimulating signals for each resonance. This may be produced by sending in multiple pairs of Stokes/anti-Stokes wavelengths and pump beams in with varying time delays between them. By varying the time delay between excitations, the molecule can be made to prefer Stokes or anti-Stokes emissions from a particular quantum state. This way, the amount of anti-Stokes radiation generated from each quantum state could be potentially probed to identify the molecule.

[147] In general, the spatial-light-modulator system of Figure 4 could be used to produce more general pulse shapes than a series of beats at a single resonance frequency. By using a more complicated pulse shape, several bonds present in a molecule can be coherently excited, and energy transferred between them in a coherent fashion. Because each molecule has some difference between the bonds presents and their relative orientation (and therefore the coupling in the quantum wave functions between them), pulses can be shaped that will selectively transfer energy between the states for a particular molecule, and not be selective for other similar molecules. In this way, the emission of stimulated Raman scattering or another coherent scattering process can be made more specific than just every molecule possessing a bond of a particular energy. With an automatically controlled pulse shaper, such as those based on spatial light modulators, feedback can be employed where the computer can test various pulse shapes, measure the resulting emitted light temporal signal using the demodulator, and progressive reshape the pulse to optimize the signal from the target molecule and exclude other molecules. Once a useful temporal field shape for stimulating a molecule has been found, it can be stored in a database and later used for identifying that molecule in the future.

[148] 2. **Reference Generator.**

[149] The reference generator takes a portion of the signal produced by the oscillator and converts it to a reference signal. This reference signal is later used to demodulate the sample signal. The reference generator is a nonlinear process that converts light in the illumination bandwidth to light in the sample's emission bandwidth, so that interference can occur between them. This nonlinear process may or may not be CARS or CSRS.

[150] A common implementation of the Reference Generator would be to focus the oscillator excitation into a sample of the same molecular species that one wishes to image. The reference signal should then be very similar to the same molecular species contained in the sample. This is because they are the same molecule, illuminated by nearly identical pulses, converting them to the output signal using resonant CARS or CSRS. The magnitude of the cross-correlation between these two signals should be great because of their similarity. In addition, if there is variability of oscillator output, having the reference generator and sample contain the same substance will respond in similar ways, so that the cross-correlation signal can remain high despite fluctuations in the oscillator. The benefit of using the same molecule in the reference generator is that it is the molecule's signal that acts as its own "fingerprint" with which the cross-correlation can use to recognize the molecule in the sample. If more selective excitation processes than CARS are used, then using the same molecule in both reference and sample will help ensure that a reference signal can be generated for a given excitation field.

[151] A nonresonant nonlinearity can be used as the reference generator as long as the peak power of the oscillator signal can excite a sufficient quantity of reference signal. High peak power can be maintained by not chirping the oscillator signal that is sent to the reference generator, while sending a relatively low peak power signal to the sample. Nonresonant CARS or CSRS can be used to generate an anti-Stokes or Stokes signal, respectively, from a medium that does not necessarily have a resonance at the frequency of the target molecule. The benefit of this is that the medium may not have to be changed each time a different molecular species is scanned for, because otherwise

a medium with a resonance at that wavelength would need to be chosen. Also, this species can act as a standard signal source against which the return signals from many samples can be compared. The nonresonant CARS can be implemented by focusing the excitation radiation into a sample of liquid that produces a CARS/CSRS signal in the same frequency band as that generated from the sample. For example, benzene will generate a CARS anti-Stokes signal in the 3000-3100 cm^{-1} frequency band.

[152] Continuum generation is another type of nonresonant nonlinear process that can be used in the reference generator. A sufficiently high peak power pulse is focused into a medium, where it excites a broad bandwidth of frequencies to be produced. If the produced frequencies overlap the emission frequency band produced in the sample, this portion of the continuum can act as a reference signal. The generated continuum must be created by a mechanism that is sufficiently stable to not be overly sensitive to fluctuations in oscillator intensity. An unstable reference signal will result in noise in the cross-correlation signal. The benefit of continuum generation is that is likely to create a broad bandwidth that will have signal in the emission bandwidth of the sample, so that the continuum need only be filtered for the needed frequency band. Also, if the oscillator employs continuum generation, it may already generate light within the emission bandwidth that can be used as a reference generator, eliminating a separate nonlinear process in the reference generator step. Some examples of materials used for continuum generation materials are optical glass, fused silica, calcium fluoride, sapphire, ethylene glycol, water, high numerical aperture optical fibers, photonic crystal optical fibers, microstructured optical fibers, dispersion-shifted optical fibers, and gas cells (e.g. cells filled with helium, argon, or nitrogen).

[153] Other candidate processes for nonresonant nonlinear reference generation include second and higher harmonic generation, stimulated Raman scattering, sum and difference frequency generation, optical parametric amplification, four-wave mixing, and self phase modulation.

[154] Figure 5 shows an implementation of a reference signal generator **500**. The concentration optics **503** are typically implemented as some combination of lenses and

mirrors. The concentration optics may also require some combination of frequency dispersive elements such as prisms, diffraction gratings, pulse shapers, and dispersive materials to prepare the temporal shape of the signal for nonlinear generation. Concentration in space and time may be necessary because the nonlinear processes are power sensitive, and depends on the strength of the nonlinear process. When the light **501** enters the nonlinear medium **505**, it undergoes conversion to a frequency band coinciding with the frequency band of the response signal from the sample. This nonlinear medium may be one of the media mentioned above, either a sample of a target molecule, a solvent, or a continuum generation medium. After exiting the nonlinear medium, the reference signal **509** is collected by reference collection relay optics **507**, where it is sent to the demodulator where it is combined with the sample signal. The collection relay optics are usually implemented as some combination of lenses and mirrors that collimate the reference field radiation. This reference field should be characterized to find its temporal structure by instruments such as Frequency Resolved Optical Gating, cross-correlation with another known signal, nonlinear sonograms, or nonlinear autocorrelations/cross-correlations.

[155] 3. **Microscope.**

[156] The microscope delivers the excitation radiation from the oscillator to the sample and collects the resulting sample emission. Inside the sample, a coherent nonlinear process such as CARS or CSRS takes place that, in the presence of a molecule of interest, will emit the sample field in response to the excitation field. The sample field is collected by the microscope and then propagated to the demodulator, where the sample field can be estimated from the measured cross-correlation between the sample field and reference field.

[157] Microscope systems can be differentiated by various implementation choices. They can either illuminate one (serial scanning) or many points (full field imaging) at a time on the tissue. If the pump and Stokes (anti-Stokes) beams of the excitation field are separated in frequency, they can be sent in either separate (non-collinear) or identical (collinear) angles into the sample. The temporal delay of the response radiation relative

to the reference may or may not be used to range molecular constituents in the tissue. The response radiation can be collected in the forward scattering (forward CARS/CSRS) or backward scattering (epi-CARS/CSRS) directions.

[158] The microscope measures spatially resolved molecular density by illuminating various points on the tissue with the oscillator field, collecting the emitted sample field, and recombining with the reference field in the demodulator. One point in the tissue may be illuminated at a given time, resulting in serial or raster scanning of the molecular density through the tissue. Alternatively, a line or a complete plane of points may be illuminated, so that data may be acquired from many points in parallel. Illuminating and measuring the radiation from an entire plane of points is called full field imaging. At the time of this writing, full field imaging is seldom used because it requires an array intensity detector such as a charge-coupled-device (CCD) to simultaneously measure the demodulated signals of all of the illuminated points. Unfortunately, as of this writing CCD arrays produce thermal dark noise at each pixel, and also have a relatively limited dynamic range of measurable intensity values. Demodulated interference signals often require very high dynamic range detection. It is conceivable that future CCD or other types of focal plane arrays (e.g. CMOS arrays) may overcome these limitations. Full field imaging also requires that the tissue be illuminated by larger amounts of power because measurable signal must be produced for an entire area rather than just one point. Since this is more likely to result in tissue damage, full field imaging will probably be used when speed of acquisition is paramount.

[159] Figure 6 shows three examples of full field CARS configurations. The "beam delivery optics" are usually implemented as some combination of mirrors and lenses that deliver beams that illuminate a wide area or line on the sample. For all of these microscope configurations, beam delivery and collection optics will typically utilize a microscope objective. The "response field collection optics" **601** are similarly implemented as a combination of lenses and mirrors that relay the response field to the demodulator so that it may be recombined with the reference and detected. The

noncollinear full field CARS **602** delivers the pump **603** and the Stokes **605** beams (assuming they are separate) at separate angles, so that the response is separated in angle from the illumination by an angle given by phase-matching conditions. The collinear geometry **607** sends the pump and Stokes radiation in the same directions (or in a single field if they can not be separated) and collects the radiation in the same direction, which can be discriminated with a dichroic beamsplitter. The epi-CARS geometry **611** collects the backscattered radiation, usually through the same objective optics that the sample is illuminated through. The epi-CARS can be discriminated from the illumination with a dichroic beamsplitter or an interference filter.

[160] If collinear CARS is used, where the pump and Stokes (anti-Stokes) beams from the excitation beams overlap, then the response field **609** will overlap the excitation beams. Then the response field frequency band should not completely overlap the excitation frequency band, so that a spectral filter may distinguish between the excitation and response fields. If non-collinear CARS is used, then the response beam can be sufficiently angularly separated (as determined by the phase-matching criterion) from the excitation radiation to be filtered by a spatial filter. However, in the case of non-phase-matched CARS, such as epi-CARS, the interaction CARS volume must be small enough so that the response is radiated effectively isotropically, so that spatial filtering is unavailable and spectral filtering should be used.

[161] The other, more commonly used alternative is serial point scanning. Serial scanning tightly focuses the oscillation signal into the tissue to create a very small volume where peak power is maximized. The focusing is usually achieved using a microscope objective. This focus is then scanned through a 3-D set of points in the tissue, and the sample signal gathered from each point is demodulated to produce a 3-D NIVI image. Since nonlinear processes are power sensitive, efficient CARS/CSRS occurs only at the focus. If the sample is small enough, the focus may be scanned through the sample by translating the sample in all three dimensions. However, it is not feasible to move large samples such as human subjects this way. In this case, the beam focus can be moved in the transverse direction by steering the beam, perhaps using galvanometer

rotated mirrors, acousto-optic modulators, or translating the lens assembly. The depth may be scanned by mechanically adjusting the distance between the lens and the tissue, perhaps using a lead-screw translator and/or a piezoelectric transducer. Since the signal can be excited at only one point at a time, one can be sure that the resulting measured sample signal is due to the interference of emissions of radiation produced in that volume only. This can be a benefit in NIVI when high phase resolution is required because one is assured that any measured phase shifts are not due to interference between molecules at disparate spatial locations. This higher phase resolution may be used to better differentiate between similar molecular species. Serial point scanning typically utilizes a single photodetector or a small number of photodetectors at the demodulator, which has the benefit that the dark current of a single photodetector is usually less than that of an entire CCD array, and a single photodetector can also typically handle a higher dynamic range of measurements.

[162] Figure 7 shows the geometry of translated serial-port scanning configurations. In all of these configurations, the sample is on a translator that moves the sample through the focus to form an image of the molecular density at various points. The translator could be a three-axis linear screw drive translator, or piezoelectric translator, or a combination of these. The beam delivery optics **701** focuses the pump **703** and Stokes (anti-Stokes) **705** beams onto the tissue at a point of interest, and the response field collection optics **707** gathers the generated anti-Stokes (Stokes) field **709** from the tissue. In the non-collinear geometry **700**, the pump and Stokes illuminate the point of interest at different angles, so that the anti-Stokes will emerge at a third angle given by phase-matching, so that the anti-Stokes is spatially separated from the illumination. In the collinear geometry **702**, the illumination and response fields emerge overlapped, so that they must be discriminated by frequency (e.g. using a dichroic beamsplitter). Finally, in the epi-CARS geometry **704**, the backscattering response fields are collected, often by the same optics through which the illumination is projected, and can be separated with a dichroic beamsplitter.

[163] Alternatively, the focus can be moved and the sample can be left stationary. This can be accomplished by a combination of tilting the illumination beams before it enters the beam delivery optics and/or translating the beam delivery optics around (Figure 8). Translating the collection optics and/or tilting the exiting response beam with a beam steering optics **801** can capture the exiting response field. The beam steering optics changes the direction of the incoming beam **803**. This can be implemented, for example, by a galvanometer-scanned rotating mirror. By changing the direction of the beam before it enters the beam delivery optics **809**, the position of the focus in the sample can be changed. The translation for the beam delivery and collection optics **807** can be implemented with a piezoelectric and/or a linear screw-drive translation stage. Translating the beam delivery and/or collection optics moves the focus with the optics through the sample. These two mechanisms can be combined to enable the three-dimensional translation of the focus through the sample. This configuration is especially convenient for the epi-CARS configuration, because the same beam steering optics and beam delivery optics **809** can be used to collect the response field **805**. For these configurations, a compensating delay may need to be incorporated into the demodulator because steering the beam and/or translating the delivery and collection optics change the travel time for the illumination and response signal through the microscope section.

[164] All of the previous scanning modes, full field imaging, translated serial-point scanning, and beam-steering serial-port scanning, use the spatial location of the illumination beam to differentiate between response signals gathered from various points in the sample. This confinement method is the same as that used by multiphoton microscopy or CARS microscopy. Other technologies such as Optical Coherence Tomography and Optical Coherence Microscopy use temporal ranging in addition to spatial confinement to further isolate the contributions of signal from various points in the sample. Because of the heterodyne nature of NIVI, this scanning mode is also available. It may be attractive for *in vivo* imaging because it will enable scanning tissue without translating the microscope objective or sample, and may scan faster because there are mechanisms for scanning the temporal delay much faster than translating

objective optics. The phase measurement capability of NIVI is useful for both temporal ranging and molecular species identification this way.

[165] Figure 9 shows two configurations of NIVI **900** and **902** utilizing temporal ranging. Temporal ranging is achieved by measuring the interference of the reference and response signals for various relative temporal delays. The temporal gating configuration superficially resembles the other epi-CARS configurations. However, unlike previous scanning modes, the depth-of-field of the focusing of the illumination from the beam delivery optics in the tissue is set to be long, because the temporal gating will be used to discriminate between molecular constituents within the depth-of-field. This temporal ranging utilizes epi-CARS/epi-CSRS because the backscattered response signal is collected from the sample by the response field connector so that the signal delay into the tissue can be timed. Because CARS/CSRS is not phase-matched for the backwards direction, the generation of a backscattered signal will only occur efficiently for small particles or edges of particles with a compatible resonance. The chief difference between this configuration and the other epi-CARS configurations is that the interference signal is scanned as a function of relative delay, and an interference signal maximum indicates the presence of a molecular species at a particular depth in the medium (corresponding to that time delay). In the full field configuration **900**, an entire plane of points is interfered with the reference signal to produce simultaneous measurements of the cross-correlation signal for the entire plane **909**. By scanning the delay, the molecular density of various planes can be measured. For the beam-steered temporal gating setup **902**, the beam is scanned through the sample by tilting the beam through the beam steering optics **903**. Beam steering provides lateral displacement of the beam, and temporal gating provides depth information so that a three-dimensional volume is scanned. The sample can be translated laterally also to scan the beam. The implementation differences between the temporal gating configurations are in the choice of depth-of-field of the illumination and collection optics, and in the demodulator. The demodulator must be designed to scan a sufficient temporal interval to capture the interference signal between the reference and response signals. This temporal interval is typically between 500 microns to 10 mm.

[166] 4. **Demodulator.**

[167] The purpose of the demodulator section is to decode the response signal from the sample so that it may be measured by relatively slow electronic equipment (slow compared to the oscillations of the electric field of the response signal). The magnitude of this demodulation signal will be related to the density of a molecular species of interest in the tissue. With knowledge of the molecular density at each point, a molecular density map, or NIVI image, can be presented to the user. This demodulation is implemented as the cross-correlation of the response signal relative to a generated reference signal. Outlined below are various optical configurations that produce this cross-correlation signal.

[168] There are various design choices that are made when choosing a demodulator. First, one needs to know whether or not full field microscopy is used. Also, one must decide whether the cross-correlation will be measured one sample at a time, or many samples in a single instant.

[169] Figure 10 shows an example of a cross-correlator that can be used to demodulate the full field CARS signal, measured for example as in Figure 6. The response field is collected and relayed by the "response field collection optics" of Figure 6 to the Response Field **1001** of Figure 10. The reference field **1003** is produced by the Reference Generator and relayed to the cross-correlator. The reference is delayed relative to the response field, and they are mutually overlapped using a beamsplitter **1005**. The "response field spectral filter" **1007** filters the recombined field for only the frequency band that contains the response field bandwidth, which removes any remaining oscillator signal. The filter can be implemented by a combination of interference filters, color glass filters, dichroic beamsplitters, or other frequency selective elements. The combined fields are imaged onto a focal plane array **1011**, such as a CCD, so that the CCD is the conjugate image plane of the sample plane. The imaging is achieved with the "relay imaging optics" **1009** which are some combination of lenses and mirrors. The intensity detected on the CCD corresponds to a single cross-correlation signal measurement collected from various points in the sample. The

"variable attenuator" **1013** is adjusted to maximize the use of the dynamic range of intensity measurements of the CCD. As the focus in the sample changes (e.g. by translating the sample), and/or the adjustable delay **1015** is changed, the entire cross-correlation signal can be measured, forming a three-dimensional data set. The typical scan range length for high numerical aperture full field imaging will be up to 100 microns. The delay can be implemented as a mirror translated by a linear-screw-drive translation stage, or by piezoelectric actuation. This same cross-correlation can also be used for temporal ranging full field NIVI, as shown in Figure 9, with the only difference being that the delay mechanism must be designed to scan a sufficiently long range of interest in the sample, typically from 500 microns to 10 mm.

[170] The measurement of a cross-correlation signal for a serial-point scanning microscope is basically the same system as for full field imaging, except that a single photodetector can be used rather than an array photodetector. The adjustable delay, response field spectral filter, and relay imaging optics can all be implemented in similar ways to the full field case. The relay imaging optics need only focus the combined response/reference signal onto the photodetector. In some cases, when only the magnitude of the cross-correlation signal at its peak needs to be measured, it may be desirable to dither the adjustable delay with a piezoelectric transducer a fraction of a wavelength, so that the magnitude of the cross-correlation signal peak can be demodulation with a multiplying mixer and low-pass filter (an electronic heterodyne demodulator). This configuration will be covered in more detail in part five. To measure the entire cross-correlation, the adjustable delay will be scanned over various time delays and the photodetector intensity signal measured.

[171] In Figures 10 and 11, a variable attenuator **1013** is used to adjust the intensity of the signal that reaches the photodetector **1101** so that its dynamic range is not exceeded. Variable attenuators can be implemented with liquid crystal shutters, neutral density filter wheels, or rotating polarizers. The spatial filter **1103** is used on the response signal to filter out spatial inhomogeneities in the response signal that could reduce the depth of modulation at the photodetector. A spatial filter would typically

consist of a telescope of two converging lenses, with a pinhole in the focal plane between the lenses to filter out power around the main diffraction focus.

[172] When the oscillator produces pulses of a low repetition rate, so that for example high peak power can be employed, one may want to measure multiple samples of the cross-correlation signal simultaneously. This can be achieved by the configurations in Figure 12. These configurations have the advantage that, with a sufficient amount of pulse power and number of photodetector array samples, the cross-correlation signal can be measured using a single pulse. Low repetition rate illumination can keep the peak power high while average power remains low.

[173] The "linear photodetector array cross-correlator" **1201** expands the response and reference signals, and interferes them with an angle between the two beams. The beam expansion is achieved by, for example, a pair of converging achromatic lenses **1203** arranged in a telescope configuration. The points on the wavefront where the two signals combine will have various time delays between them. A cylindrical lens **1205** then focuses the beams into a line image on a linear photodiode array **1207**. This concentrates the signal to adjust for the narrow height of the detector pixels. Each intensity sample on the linear photodetector array indicates a sample of cross-correlation of the two fields with various relative time delays, with a constant intensity signal added. The recorder would then read the linear CCD signal so that a complete NIVI image can be assembled.

[174] The primary disadvantage of this system is that the modulation of the intensity signal on the linear photodetector array is rather low, and the dynamic range of typical CCDs is likewise low, so that the signal cannot be measured to a high signal-to-noise ratio. To combat this, one can utilize the Fourier-transform cross-correlator **1209**. Rather than directly measuring samples of the cross-correlation, it interferes the two signals and measures their spectral decomposition. The beamsplitter **1005** combines the response and reference signals, which are filtered for the response bandwidth by utilizing a response field spectral filter **1007**. This signal is then filtered by a frequency dispersive element **1211** such as a diffraction grating that scatters each frequency to a

different angle. The power of each frequency is then focused to a different pixel on the linear photodetector array **1215** by using a focusing element **1213** (typically a combination of lenses or mirrors). The samples of the intensity measured on the photodetector indicate the real part of the Fourier transform of the cross-correlation signal. The recorder may compute an inverse discrete Fourier transform of the linear photodetector array intensity samples to recover the cross-correlation. The Fourier transform already performs the necessary Hilbert transform to infer the imaginary part of the cross-correlation signal. However, because it is an intensity measurement, the cross-correlation signal, a time-reversed version, and the autocorrelation of the response signal are superimposed in the intensity signal. By choosing the adjustable delay longer than the length of the cross-correlation temporal signal, the reconstruction of these signal components will not overlap in the time domain, and so the recorder device may distinguish the cross-correlation from its mirror image and the autocorrelation.

[175] The adjustable delay may also be dithered a small amount to introduce a phase shift into each measured frequency component at the linear photodetector array. By utilizing at least three linearly independent phase shifts, the phase of each Fourier component can be known and the time domain cross-correlation computed using an inverse discrete Fourier transform. However, then the ability to measure the cross-correlation of a single pulse will be lost. Another possibility is to use a linear photodetector array with several rows (at least three) rather than just one row of pixels. By tilting one of the wavefronts slightly vertically with respect to the other, a small phase shift can be introduced in the measured cross-correlation signal between rows on the linear photodetector. With a sufficiently large phase shift, the complex amplitude of each Fourier component can be inferred from the intensity samples from each column on the linear photodetector by utilizing the discrete Fourier transform. Since all of the time delayed cross-correlation samples are measured simultaneously, this is a single-pulse measurement.

[176] To minimize the effect of dark charge built up in the linear CCD detector, it is best to discard the charge as soon as possible before the pulse arrives, and read the

CCD as soon as possible after the pulse arrives. The speed of readout should be as fast as possible given limitations in accuracy due to the readout noise.

[177] 5. **Recorder.**

[178] The recorder accumulates the samples of the cross-correlation signal gathered from various points in the sample, processes this data, and presents these as a human-interpretable image. It is usually implemented as a data acquisition digital computer with some method of automatic control of the adjustable mechanisms (such as delay lines, piezoelectric actuators, and galvanometer mirrors), analog-to-digital conversion, and some form of image output device such as a screen or printer. It may also have control of the oscillator itself, to automatically tune the wavelengths or bandwidths of output or control the rate or timing of pulse output. It may also control the delay lines or spatial light modulators in the pulse shaping mechanisms of Figure 4.

[179] Typically, the recorder will scan the illumination through the sample and/or the adjustable time delay and measure the cross-correlation signal. Unless the Fourier-Transform cross-correlator is used, the intensity samples represent the cross-correlation signal with a constant level added that could be discarded. The magnitude of the cross-correlation signal indicates the presence of the target molecular species. For CARS/CSRS the magnitude of the cross-correlation signal is related to the second power of the molecular density of that species. By demodulating the magnitude of the cross-correlation signal, the molecular density can be estimated for the points on the cross-correlation signal for the areas from which the response signal was collected to form that cross-correlation signal. Both spatial confinement and temporal ranging can be used to differentiate between the signals due to molecular densities at different locations in the sample.

[180] Figure 13 documents the relationship between the elements of the recorder **1300**. For serial-scanning or temporal-scanning configurations, the instructions from the operator are entered via a human interface device **1304** into the digital computer **1301**. The digital computer controls the scanned beam position in the sample using the

galvanometer mirror angles of the galvanometer scanned mirrors **1303**, the oscillator **1302**, the adjustable delay line **1309**, the position of the microscope objective position **1305**, and/or the position of the sample translator **1307** to scan the illumination through the sample. In general only a subset of these need be controlled to scan the three-dimensional volume of the tissue. If full field imaging is used, usually only the depth and/or the delay line need be scanned. The cross-correlation signal as measured by the intensity is converted to a voltage by the photodetector **1311**, which is in turn converted to a digital sample by the analog-to-digital converter **1313**. The digital computer varies the delay line and/or reads various pixels from the photodetector (if it is an array detector) to determine the cross-correlation signal. If a Fourier-transform cross-correlator is used, the computer will need to compute the inverse discrete Fourier transform of the signal to find its cross-correlation from the intensity samples. The computer then associates the magnitude of the cross-correlation signal with the molecular density, and assembles these magnitudes into a density map of the sample. This density map is stored in the storage device **1312** presented on the visual display **1315** and/or made into a physical representation with a printer or stereolithography device **1317**.

[181] To aid in the measurement of the magnitude of the cross-correlation signal, the configuration **1400** of Figure 14 is suggested.

[182] Because the digital computer **1301** can record only relatively slow signals, a dither oscillator **1401** may be used to add a small high-frequency dither signal into the adjustable delay line **1309** that produces a periodic perturbation of the delay in the signal of a magnitude usually of less than one wavelength. This same signal is multiplied by the received photodetector **1403** interference signal to demodulate it, and is low pass filtered through the low pass filters **1405** to remove harmonics of the dither frequency. Both the dither and its quadrature signal are demodulated, because these correspond to the real and imaginary parts of the complex cross-correlation amplitude. The analog-to-digital converter **1313** will then be directly measuring a quantity proportional to the complex magnitude of the cross-correlation for the delay line

position. The computer can utilize the digitized real and imaginary cross-correlation components to display the amplitude and phase of the cross-correlation signal. The amplitude will correspond to the magnitude of the reflection, and the phase will correspond to the Doppler shift of the reflection. If a very fast dither signal is desired (above 10 kHz), an electro-optic modulator or acoustical-optic modulator can be placed in the adjustable delay line system to rapidly modulate the delay a small amount.

[183] The magnitude of the cross-correlation signal depends both on the density of molecules available to produce the response signal, and on the temporal structure of the signals radiated from the molecules. The temporal signal produced by the molecules can be inferred from the cross-correlation signal. If $f(t)$ is the temporal response signal produced by a molecule, and $g(t)$ is the known reference signal, the measured cross-correlation $\Gamma(\tau)$ is given by:

[184] $$\Gamma(\tau) = \int_{-\infty}^{\infty} f(t)g(t-\tau)dt$$

[185] If $\tilde{F}(\omega)$, $\tilde{G}(\omega)$, and $\tilde{\Gamma}(\omega)$ are the Fourier transforms of $f(t)$, $g(t)$, and $\Gamma(\tau)$ respectively, then $\tilde{\Gamma}(\omega) = \tilde{F}(\omega)\tilde{G}(-\omega)$. The function $\tilde{\Gamma}(\omega)$ can be computed from the Fourier transform of the cross-correlation and $\tilde{G}(\omega)$ can be computed from the measured reference signal. The Fourier transform $\tilde{F}(\omega)$ of temporal signal $f(t)$ can then be estimated by $\tilde{F}(\omega) = \tilde{\Gamma}(\omega)\tilde{G}(-\omega)^* / \left(\tilde{G}(-\omega)^2 + N(\omega)^2\right)$ (change to equation), where $N(\omega)$ is an estimated power-spectral-density of the noise. An inverse Fourier transform of $\tilde{F}(\omega)$ yields $f(t)$.

[186] The recorder can perform this computation and thus recover $f(t)$ for various molecular species. An unknown molecule may be identified by comparing its temporal signal to known molecular signals. In addition, by looking at the frequencies ω of phase $\arg \tilde{F}(\omega)$ where the phase changed rapidly, the resonance frequencies of the molecules of interest can be determined to high precision. A library of the temporal

responses and resonance frequencies of various molecules for CARS excitation may be built up and used to identify unknown molecules *in vivo*. We also note that the same cross-correlation measurement configurations can be used to measure the OCT backscattered signal due to the scattering of the excitation radiation. To do this, the excitation and response bands should be separated with a dichroic beamsplitter, and the cross-correlation setups implemented separately to eliminate photon noise from the excitation band being measured in the response band. Different cross-correlation setups may be used for each configuration, e.g. a single-pulse measurement system for the conventional excitation and a time delay cross-correlator as in Figure 11.

[187] ## EXAMPLES

[188] Theory of CARS interferometry and inference

[189] We consider a focused optical pulse, with an electric field at the focus given in the frequency domain by $\tilde{E}_i(\omega)$. At the focus there is a medium with a complex Raman spectrum given by $\chi^{(3)}(\Omega)$ where Ω is the Raman frequency. The anti-Stokes/Stokes electric field produced by the CARS/CSRS (Coherent Stokes Raman Scattering) process at the focus is given by:

[190] $$P^{(3)}(\Omega) = \chi^{(3)}(\Omega)\int_0^\infty \tilde{E}_i(\omega+\Omega)\tilde{E}_i(\omega)^* d\omega \qquad (1)$$

[191] $$\tilde{E}_A(\omega) = \int_0^\omega P^{(3)}(\Omega)\tilde{E}_i(\omega-\Omega)d\Omega \qquad \text{(CARS)} \qquad (2)$$

[192] $$\tilde{E}_S(\omega) = \int_0^\omega P^{(3)}(\Omega)^* \tilde{E}_i(\omega+\Omega)d\Omega \qquad \text{(CSRS)} \qquad (3)$$

[193] These equations assume that classical electromagnetic field approximations hold, a perturbative interaction with the sample, and that there are no levels directly resonant with any of the individual frequencies in the pulse. These equations give the time

evolution of a CARS/CSRS process involving many possible simultaneous Raman-active vibrations. In CARS, the molecule is excited by two frequencies ω_1 and ω_2 that are separated by the Raman vibrational frequency $\omega_1 - \omega_2 = \Omega$. The upper frequency ω_1 is called the pump, and the lower frequency ω_2 the Stokes. In general, the signal consists of not just two discrete frequencies but a continuous spectrum. In this case, Eq. 1 represents the excitation of the molecule through SRS. A vibration at frequency Ω is excited by every pair of optical frequencies with a frequency difference of Ω. The excitation at any given Raman frequency is given by the polarization $P^{(3)}(\Omega)$. This excitation is converted to anti-Stokes radiation $\tilde{E}_A(\omega)$ by a second SRS process whereby the pump signal at frequency ω_1 is partially converted to the anti-Stokes signal at frequency $\omega_A = \omega_1 + \Omega = 2\omega_1 - \omega_2$. For broadband signals, the anti-Stokes signal must be summed over all pump frequencies, and so Eq. 2 results. The CSRS process is likewise modeled by Eq. 1 and Eq. 3. In general, both CARS and CSRS are generated simultaneously for any given pulse.

[194] The information about the sample is contained in its Raman spectrum $\chi^{(3)}(\Omega)$. By sending in a pulse with a known electric field $\tilde{E}_i(\omega)$, and measuring the returned anti-Stokes signal $\tilde{E}_A(\omega)$, we would like to infer $\chi^{(3)}(\Omega)$. Unfortunately, at optical frequencies the electric field $\tilde{E}_A(\omega)$ is not directly measurable because it oscillates on a time scale too fast to be electronically demodulated, necessitating interferometric demodulation.

[195] To demodulate the anti-Stokes field, we generate a reference field $\tilde{R}(\omega)$ from the oscillator field that contains frequencies same as the anti-Stokes radiation. The interferometric demodulator measures the intensity of the cross-correlation of the anti-Stokes and reference fields, which is given by the frequency spectrum $\tilde{I}(\omega) = \tilde{R}(\omega)^* \tilde{E}_A(\omega)$. A good method of interferometric demodulation for the techniques disclosed is spectral interferometry [1]. Spectral interferometry may be especially attractive because it allows the signal to be sampled in one shot to minimize transient

effects, and can achieve higher signal-to-noise ratio than conventional temporal interferometry. Putting the steps of generating the anti-Stokes radiation and interferometry together, the cross-correlation spectrum is given by:

[196] $$\tilde{I}(\omega) = \tilde{R}(\omega)^* \int_0^{\omega} \chi^{(3)}(\Omega)\tilde{E}_i(\omega-\Omega)\int_0^{\infty}\tilde{E}_i(\omega'+\Omega)\tilde{E}_i(\omega')^* d\omega' d\Omega \quad \text{(CARS)} \qquad (4)$$

[197] $$\tilde{I}(\omega) = \tilde{R}(\omega)^* \int_0^{\omega} \chi^{(3)}(\Omega)^* \tilde{E}_i(\omega+\Omega)\int_0^{\infty}\tilde{E}_i(\omega'+\Omega)^* \tilde{E}_i(\omega') d\omega' d\Omega \quad \text{(CSRS)} \qquad (5)$$

[198] An important consequence of Eqs. 4 and 5 is that the cross-correlation spectrum $\tilde{I}(\omega)$ is a linear function of the Raman spectrum $\chi^{(3)}(\Omega)$. Because of this, linear estimation can be used to estimate $\chi^{(3)}(\Omega)$ from $\tilde{I}(\omega)$. To better see this, we reform Eqs. 4 and 5 as linear operators $\hat{\mathbf{A}}$ and $\hat{S}$ with kernels $A(\omega,\Omega)$ and $S(\omega,\Omega)$ respectively:

[199] $$\tilde{I}(\omega) = \hat{\mathbf{A}}\chi = \int_0^{\omega} \chi^{(3)}(\Omega) A(\omega,\Omega) d\Omega \quad \text{where}$$

[200] $$A(\omega,\Omega) = \tilde{R}(\omega)^* \tilde{E}_i(\omega'-\Omega)\int_0^{\infty}\tilde{E}_i(\omega'+\Omega)\tilde{E}_i(\omega')^* d\omega' \qquad \text{(CARS)} \quad (6)$$

[201] $$\tilde{I}(\omega) = \hat{S}\chi^* = \int_0^{\omega} \chi^{(3)}(\Omega)^* S(\omega,\Omega) d\Omega \quad \text{where}$$

[202] $$S(\omega,\Omega) = \tilde{R}(\omega)^* \tilde{E}_i(\omega'+\Omega)\int_0^{\infty}\tilde{E}_i(\omega'+\Omega)^* \tilde{E}_i(\omega') d\omega' \qquad \text{(CSRS)} \quad (7)$$

[203] With the CARS interferometry process now expressed as a linear operator $\hat{\mathbf{A}}$ mapping $\chi^{(3)}(\Omega)$ to $\tilde{I}(\omega)$, an inverse operator can be found to estimate $\chi^{(3)}(\Omega)$ from $\tilde{I}(\omega)$. Many inverse operators are possible and should be chosen on the basis of stability, ability to model the physical system and noise, and ability to incorporate confidence measures of the data. We present an inverse operator based on the weighted Tikhonov regularized least-squares solution. The least-squares solution by

itself attempts to find a solution for $\chi^{(3)}(\Omega)$ that minimizes the Euclidean distance $\left\| I - \hat{A}\chi \right\|$ for the linear system $I = \hat{A}\chi$. Unfortunately, this in practice tends to produce poor results due to poor conditioning and the inability to incorporate confidence information about the data. A solution is to augment the linear system with a weighting operator

[204] $$\hat{W}I = \tilde{W}(\omega)\tilde{I}(\omega) \qquad (8)$$

[205] where $\tilde{W}(\omega)$ is a weighting function that is of unit value inside the anti-Stokes bandwidth and zero outside. The weighted linear system becomes $\hat{W}I = \hat{W}\hat{A}\chi$. The Tikhonov regularized solution to this system is $\chi = \left(\hat{A}^H\hat{W}^H\hat{W}\hat{A} + \varepsilon\hat{I}\right)^{-1}\hat{A}^H\hat{W}^H\hat{W}I$ where the operator $\hat{I}$ is the identity operator (for CSRS it is $\chi = \left(\hat{S}^H\hat{W}^H\hat{W}\hat{S} + \varepsilon\hat{I}\right)^{-1}\hat{S}^H\hat{W}^H\hat{W}I$). The constant $\varepsilon > 0$ is the set to account for the magnitude of noise in the measurement, e.g. from thermal or photon noise. In practice, this operator can be computed numerically using iterative linear solution methods such as the preconditioned conjugate gradient method.

[206] To perform this inverse in practice, we used the preconditioned conjugate gradient method. This method is a method of computing the inverse of a linear system for a particular solution without explicitly computing the inverse of the matrix. Such "sparse matrix" methods are used because the explicit inverse would be intractable or far more memory and computationally intensive than the sparse solution. To make conjugate gradient methods work quickly, a preconditioner must be selected that approximates the inverse solution to the true problem. While a solution can be obtained if no preconditioner is used (the identity preconditioner), the conjugate gradient method is greatly enhanced by the choice of preconditioner. The conjugate gradient algorithm is well-known in the literature and so we just review its application to the inverse of this problem.

[207] The problem we solve is $\left(\hat{A}^H\hat{W}^H\hat{W}\hat{A}+\varepsilon\hat{I}\right)\chi=\hat{A}^H\hat{W}^H\hat{W}I$ for the CARS process and $\left(\hat{S}^H\hat{W}^H\hat{W}\hat{S}+\varepsilon\hat{I}\right)\chi=\hat{S}^H\hat{W}^H\hat{W}I$ for the CSRS process. These are total-least-squares solutions with Tikhonov regularization and weighting of the anti-Stokes (Stokes) spectrum. While the total-least-squares solution in general is not simply computable using simple operators such as the discrete Fourier transform, we can find an approximate inverse that is. We precompute $\chi'=\hat{A}^H\hat{W}^H\hat{W}I$ and call the operator $\hat{B}=\hat{A}^H\hat{W}^H\hat{W}\hat{A}+\varepsilon\hat{I}$ so that this system becomes $\hat{B}\chi=\chi'$ for CARS. (For CSRS $\chi'=\hat{S}^H\hat{W}^H\hat{W}I$ and $\hat{B}=\hat{S}^H\hat{W}^H\hat{W}\hat{S}+\varepsilon\hat{I}$). We assume that the anti-Stokes (Stokes) and Raman spectra are sampled to discrete frequencies. The matrix $\hat{W}$ is a weighting operator implemented by multiplying each anti-Stokes (Stokes) frequency by its corresponding weight. To speed up the solution, rapid ways of performing the operators

[208] $\hat{A}^H$, $\hat{A}$, $\hat{S}^H$, and $\hat{S}$ must be found. Fortunately, these can be computed using the Fast Fourier transform. To do this, we assume that the vector E is a sampled version of the incoming electric field $\tilde{E}_i(\omega)$ in the frequency domain. Likewise, we assume that R is a frequency sampled version of the reference pulse $\tilde{R}(\omega)$. We use the operator DFT to denote the discrete Fourier transform operation and DFT^{-1} to indicate its inverse, which is typically implemented using the Fast Fourier Transform algorithm.

[209] To compute $\hat{A}$, we find the power spectrum of the instantaneous intensity of the signal $I_E=DFT\left[\left|DFT^{-1}[E]\right|^2\right]$. The operator $\hat{A}\chi=(R^*)DFT\left[DFT^{-1}[E]DFT^{-1}[I_E\chi]\right]$. The discrete Fourier transform here is used as a rapid way to implement the convolution and product operations of Eq 6. Likewise, $\hat{S}\chi=(R^*)DFT\left[DFT^{-1}[E]DFT^{-1}[I_E\chi]^*\right]$ (notice the complex conjugate operator). Written this way, the Hermitian adjoints are easy to find. $\hat{A}^HI=I_E^*DFT\left[DFT^{-1}[E]^*DFT^{-1}[RI]\right]$ and $\hat{S}^HI=I_EDFT\left[DFT^{-1}[E]^*DFT^{-1}[RI]\right]^*$. To save additional computation, the real-to-complex DFT can be used.

[210] Based on the forward operators for $\hat{A}$ and $\hat{S}$, we can approximate an inverse for these operators that can be used as a preconditioner. These are based on a least-squares solution to each step of the forward operator, rather than computing the least-squares solution to the total operator. The approximate inverse used for $\hat{A}$ and $\hat{S}$ are:

[211] $$\hat{A}^{-1}_{approx} I = \frac{I_E^{\ *}}{|I_E|^2 + I_{E0}^{\ 2}} DFT\left[\frac{DFT^{-1}\left[\frac{RI}{|R|^2+I_0^{\ 2}}\right] DFT^{-1}[E]^*}{|DFT^{-1}[E]|^2+E_0^{\ 2}}\right] \quad (9)$$

[212] $$\hat{S}^{-1}_{approx} I = \frac{I_E}{|I_E|^2 + I_{E0}^{\ 2}} DFT\left[\frac{DFT^{-1}\left[\frac{RI}{|R|^2+I_0^{\ 2}}\right] DFT^{-1}[E]^*}{|DFT^{-1}[E]|^2+E_0^{\ 2}}\right]^* \quad (10)$$

[213] These are approximate inverses that have been regularized by three constants: I_0^2, E_0^2, and I_{E0}^2. These constants appear in divisors to prevent a division by a small magnitude number, which would produce instability in the solution. The constants are usually chosen based on the maximum value of the magnitude of the corresponding quantity. The constant I_0^2 can be chosen based on the maximum magnitude squared of a frequency component in the measured signal vector I, typically some small fraction of it (e.g. 0.01). The constant E_0^2 is chosen by the maximum magnitude squared of a temporal sample of the sample field $DFT^{-1}[E]$ (because the corresponding operation occurs in the time domain rather than the frequency domain). Finally, the constant I_{E0}^2 is chosen by the maximum magnitude squared of any frequency sample of I_E.

[214] A summary of how to apply the preconditioned conjugate gradient algorithm to the inverse problem is given in Figure 16.

[215] The method works by taking an initial random guess for the Raman spectrum that can be a random vector, a vector of constant values, or perhaps a prcondictioned conjugate gradient. From this vector the residual is calculated and reduces each

iteration, with the two conjugate search directions updated. For this algorithm to work with CSRS rather than CARS, substitute $\hat{S}, \hat{S}^H$, $\hat{S}^{-1}_{approx}$ for $\hat{A}$, $\hat{A}^H$, and $\hat{A}^{-1}_{approx}$ in Figure 16.

[216] Sometimes to improve accuracy, a conjugate gradient "restart" will be needed. To do this, the best solution from the last series of conjugate gradient iterations can be used as the starting point for the next series rather than a random guess. The best solution is found when the solution $\chi^{(i)}$ appears to converge and stabilize near the solution, and then diverges or oscillates around the solution. Usually only one restart is needed.

[217] Stimulation of a band of Raman frequencies

[218] As an example of a pulse combination that can stimulate a band of Raman frequencies [2], we present a simulation of this technique using experimentally realistic values. We assume that the laser source is an ultrabroadband Ti-sapphire laser producing transform-limited pulses of uniform power spectral density between 700-1000 nm. The bandwidth from 800-1000 nm is reserved for stimulating CARS, while the bandwidth from 700-800 nm is used as the reference pulse. A hypothetical molecule was used as the sample, with Lorentzian resonances at 800 cm^{-1}, 900 cm^{-1}, 1000 cm^{-1}, and 1100 cm^{-1}. These are frequencies in the Raman fingerprint region which is useful for practical molecular identification. A pulse with $\Omega_L = 700$ cm^{-1}, $\Omega_H = 1300$ cm^{-1}, and $T = 5$ ps was used in Eq. 33 as the hypothetical stimulation pulse. The Ω_L and Ω_H are chosen to span the range of desired measured Raman frequencies. In practice, there should be extra measured bandwidth around the desired range (typically 20-25%) to ensure that the desired frequencies are sufficiently sampled. The time T is chosen to be on the order of the lifetime of the resonances to maximize resolution and signal.

[219] The results of the simulation are presented in Figure 25. Part (a) shows the magnitude of the illumination pulse, including the overlapped region where the Raman

excitation occurs. Part (b) shows the spectrum of the illumination pulse. Part (c) shows the spectrum of the generated anti-Stokes light, simulated using Eqs. 1 and 2. Only the portion of the spectrum shorter than 800 nm is available for the inverse, because anti-Stokes in the same frequency band as the illumination is inseparable. Part (d) is the beat frequency spectrum, or the Fourier transform of the intensity of the pulse. This is the possible Raman frequencies the pulse can stimulate. Part (e) is the magnitude of the Raman susceptibility, and part (f) is the recovered Raman susceptibility using the regularized least-squares solution.

[220] A second simulation demonstrates the ability to recover a small portion of the Raman spectrum in detail. In this simulation, we desire to simulate the measurement of the relative amounts of DNA (deoxyribonucleic acid) which has a resonance at 1094 cm^{-1}, and RNA (ribonucleic acid) which has a resonance at 1101 cm^{-1}. A hypothetical Raman spectrum was created which has Lorentzian resonances at both frequencies. A pulse is designed such that $\Omega_L = 1070$ cm^{-1}, $\Omega_H = 1130$ cm^{-1}, and $T = 5$ ps. The simulation of the anti-Stokes signal and the reconstruction of the Raman spectrum are shown in Figure 26. Because of the relatively short pulse interval, there is a noticeable resolution loss to the spectrum, but the resonance lines are distinct. The short dephasing time in liquids, especially water is the practical limit on the instrument resolution *in vivo*.

[221] # REFERENCES

[222] 1. S.A. Boppart and D.L. Marks, U.S. patent application serial no. 10/717,437 "NONLINEAR INTERFEROMETRIC VIBRATIONAL IMAGING", the entire contents of which are hereby incorporated by reference (filed 11 Nov. 2003).

[223] 2. D.L. Marks, S.A. Boppart, *Nonlinear interferometric vibrational imaging* (2003), E-print@arxiv.org/physics/0311071, URL http://www.arxiv.org/abs/physics/0311071)

[224] 3. D.L. Marks, C. Vinegoni, J.S. Bredfeldt, and S.A. Boppart, *Pulse shaping strategies for nonlinear interferometric vibrational imaging optimized for biomolecular imaging* Conference Proceeding: EMBC 2004: 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (1-5 Sept. 2004, San Francisco, CA), vol. 7, pp. 5300 (accession number 8255487).

[225] 4. Schmitt JM, Knuttel A, Bonner RF. Measurements of optical properties of biological tissues by low-coherence reflectometry. Appl. Opt. 32:6032-6042, 1993.

[226] 5. Sergeev AM, Gelikonov VM, Gelikonov GV, Feldchtein FI, Kuranov RV, Gladkova ND. *In vivo* endoscopic OCT imaging of precancer and cancer states of human mucosa. Opt. Express 1:432-440, 1997.

[227] 6. Tearney GJ, Brezinski ME, Bouma BE, Boppart SA, Pitris C, Southern JF, Fujimoto JG. *In vivo* endoscopic optical biopsy with optical coherence tomography. Science. 276:2037-2039, 1997.

[228] 7. Boppart SA, Bouma BE, Pitris C, Tearney GJ, Southern JF, Brezinski ME, Fujimoto JG. Intraoperative assessment of microsurgery with three-dimensional optical coherence tomography. Radiology. 208:81-86, 1998.

[229] 8. Hee MR, Izatt JA, Swanson EA, Huang D, Schuman JS, Lin CP, Puliafito CA, Fujimoto JG. Optical coherence tomography of the human retina. Arch. Ophthalmol. 113:325-332, 1995.

[230] 9. Puliafito CA, Hee MR, Lin CP, Reichel E, Schuman JS, Duker JS, Izatt JA, Swanson EA, Fujimoto JG. Imaging of macular disease with optical coherence tomography (OCT). Ophthalmology 102:217-229, 1995.

[231] 10. Puliafito CA, Hee MR, Schuman JS, Fujimoto JG. *Optical Coherence Tomography of Ocular Diseases*. Slack, Inc, Thorofare, NJ, 1995.

[232] 11. Schmitt JM, Knuttel A, Yadlowsky M, Eckhaus AA. Optical coherence tomography of a dense tissue: statistics of attenuation and backscattering. Phys. Med. Biol. 39:1705-1720, 1994.

[233] 12. Schmitt JM, Yadlowsky MJ, Bonner RF. Subsurface imaging of living skin with optical coherence microscopy. Dermatology 191:93-98, 1995.

[234] 13. Sergeev AM, Gelikonov VM, Gelikonov GV, Feldchtein FI, Kuranov RV, Gladkova ND, Shakhova NM, Snopova LB, Shakov AV, Kuznetzova IA, Denisenko AN, Pochinko VV, Chumakov YP, Streltzova OS. *In vivo* endoscopic OCT imaging of precancer and cancer states of human mucosa. Opt Express 1:432-440, 1997.

[235] 14. Profio AE, Doiron DR. Transport of light in tissue in photodynamic therapy of cancer. Photochem. Photobiol. 46:591-599, 1987.

[236] 15. Tearney GJ, Brezinski ME, Boppart SA, Bouma BE, Weissman N, Southern JF, Swanson EA, Fujimoto JG. Catheter-based optical imaging of a human coronary artery. Circulation 94:3013, 1996.

[237] 16. Tearney GJ, Brezinski ME, Southern JF, Bouma BE, Boppart SA, Fujimoto JG. Optical biopsy in human gastrointestinal tissue using optical coherence tomography. Amer. J. Gastroenterol. 92:1800-1804, 1997.

[238] 17. Tearney GJ, Brezinski ME, Southern JF, Bouma BE, Boppart SA, Fujimoto JG. Optical biopsy in human urologic tissue using optical coherence tomography. J. Urol. 157:1915-1919, 1997.

[239] 18. Boppart SA, Brezinski ME, Pitris C, Fujimoto JG. Optical Coherence Tomography for Neurosurgical Imaging of Human Intracortical Melanoma. Neurosurgery 43:834-841, 1998.

[240] 19. Bouma BE, Tearney GJ, Boppart SA, Hee MR, Brezinski ME, Fujimoto JG. High resolution optical coherence tomographic imaging using a modelocked Ti:Al_2O_3 laser. Opt. Lett. 20:1486-1488, 1995.

[241] 20. Drexler W, Morgner U, Kartner FX, Pitris C, Boppart SA, Li X, Ippen EP, Fujimoto JG. *In vivo* ultrahigh resolution optical coherence tomography. Opt. Lett. 24:1221-1223, 1999.

[242] 21. Tearney GJ, Bouma BE, Boppart SA, Golubovic B, Swanson EA, Fujimoto JG. Rapid acquisition of *in vivo* biological images using optical coherence tomography. Opt. Lett. 21:1408-1410, 1996.

[243] 22. Chen Z, Milner TE, Srinivas S, Wang X. Noninvasive imaging of *in vivo* blood flow velocity using optical Doppler tomography. Opt. Lett. 22:1119-1121, 1997.

[244] 23. Yazdanfar S, Kulkarni MD, Izatt JA. High resolution imaging of *in vivo* cardiac dynamics using color Doppler optical coherence tomography. Opt. Express. 1:424-431, 1997.

[245] 24. de Boer JF, Milner TE, van Germert MJC, Nelson SJ. Two-dimensional birefringence imaging in biological tissue by polarization sensitive optical coherence tomography. Opt. Lett. 22:934-936, 1997.

[246] 25. Tearney GJ, Boppart SA, Bouma BE, Brezinski ME, Weissman NJ, Southern JF, Fujimoto JG. Scanning single-mode fiber optic catheter-endoscope for optical coherence tomography. Opt. Lett. 21:1-3, 1996.

[247] 26. Boppart SA, Bouma BE, Pitris C, Tearney GJ, Fujimoto JG. Forward-imaging instruments for optical coherence tomography. Opt. Lett. 22:1618-1620, 1997.

[248] 27. Tearney GJ, Brezinski ME, Bouma BE, Boppart SA, Pitris C, Southern JF, Fujimoto JG. *In vivo* endoscopic optical biopsy with optical coherence tomography. Science 276:2037-2039, 1997.

[249] 28. Bouma BE, Tearney GJ, Compton CC, Nishioka NS. High-resolution imaging of the human esophagus and stomach *in vivo* using optical coherence tomography. Gastrointest. Endosc. 51:467-474, 2000.

[250] 29. Sivak MV Jr, Kobayashi K, Izatt JA, Rollins AM, Ung-Runyawee R, Chak A, Wong RC, Isenberg GA, Willis J. High-resolution endoscopic imaging of the gastrointestinal tract using optical coherence tomography. Gastrointest. Endosc. 51:474-479, 2000.

[251] 30. Li X, Boppart SA, Van Dam J, Mashimo H, Mutinga M, Drexler W, Klein M, Pitris C, Krinsky ML, Brezinski ME, Fujimoto JG. Optical coherence tomography: advanced technology for the endoscopic imaging of Barrett's esophagus. Endoscopy 32:921-930, 2000.

[252] 31. Boppart SA, Brezinski ME, Bouma BE, Tearney GJ, Fujimoto JG. Investigation of developing embryonic morphology using optical coherence tomography. Dev. Biol. 177:54-63, 1996.

[253] 32. Boppart SA, Brezinski ME, Tearney GJ, Bouma BE, Fujimoto JG. Imaging developing neural morphology using optical coherence tomography. J. Neurosci. Meth. 2112:65-72, 1996.

[254] 33. Boppart SA, Tearney GJ, Bouma BE, Southern JF, Brezinski ME, Fujimoto JG. Noninvasive assessment of the developing *Xenopus* cardiovascular system using optical coherence tomography. Proc. Natl. Acad. Sci. USA 94:4256-4261, 1997.

[255] 34. Boppart SA, Bouma BE, Pitris C, Southern JF, Brezinski ME, Fujimoto JG. *In vivo* cellular optical coherence tomography imaging. Nature Med. 4:861-864, 1998.

[256] 35. Pitris C, Goodman AK, Boppart SA, Libus JJ, Fujimoto JG, Brezinski ME. High resolution imaging of gynecological neoplasms using optical coherence tomography. Obstet. Gynecol. 93:135-139, 1999.

[257] 36. Pitris C, Jesser C, Boppart SA, Stamper D, Brezinski ME, Fujimoto JG. Feasibility of optical coherence tomography for high resolution imaging of human gastrointestinal tract malignancies. J. Gastroenterol. 35:87-92, 1999.

[258] 37. Boppart SA. *Surgical Diagnostics, Guidance, and Intervention using Optical Coherence Tomography*. Ph.D. Thesis. Harvard-MIT Division of Health Sciences and Technology, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, 1998.

[259] 38. Huang D, Swanson EA, Lin CP, Schuman JS, Stinson WG, Chang W, Hee MR, Flotte T, Gregory K, Puliafito CA, Fujimoto JG. Optical Coherence Tomography. Science 254:1178-1181, 1991.

[260] 39. Fujimoto JG, Brezinski ME, Tearney GJ, Boppart SA, Bouma BE, Hee MR, Southern JF, Swanson EA. Biomedical imaging and optical biopsy using optical coherence tomography. Nature Medicine 1:970-972, 1995.

[261] 40. Brezinski ME, Tearney GJ, Bouma BE, Izatt JA, Hee MR, Swanson EA, Southern JF, Fujimoto JG. Optical coherence tomography for optical biopsy: properties and demonstration of vascular pathology. Circulation 93:1206-1213, 1996.

What is claimed is:

1. A method of examining a sample, comprising:

exposing of the sample to a pump pulse of electromagnetic radiation for a first period of time,

exposing of the sample to a stimulant pulse of electromagnetic radiation for a second period of time which overlaps in time with at least a portion of the first exposing, to produce a signal pulse of electromagnetic radiation for a third period of time, where the first and third periods of time are each greater than the second period of time, and

interfering the signal pulse with a reference pulse of electromagnetic radiation, to determine which portions of the signal pulse were produced during the exposing of the sample to the stimulant pulse.

2. The method of any of the preceding claims, further comprising identifying presence of molecules in the sample from a portion of the signal pulse not produced during the exposing of the sample to the stimulant pulse.

3. The method of any of the preceding claims, wherein the portion of the signal pulse not produced during the exposing of the sample to the stimulant pulse comprises Stokes photons.

4. The method of any of the preceding claims, wherein the portion of the signal pulse not produced during the exposing of the sample to the stimulant pulse comprises anti-Stokes photons.

5. The method of any of the preceding claims, wherein the sample is selected from the group consisting of a tissue sample, a single cell, and a patient.

6. The method of any of the preceding claims, wherein the sample is selected from the group consisting of a tissue sample, a single cell, and a patient.

7. The method of any of the preceding claims, further comprising producing the reference pulse by exposing a reference material to a fourth pulse of electromagnetic radiation.

8. The method of any of the preceding claims, wherein the first period of time is at least twice as long as the second period of time.

9. The method of any of the preceding claims, wherein the presence of the molecules is indicative of a disease state in a patient.

10. The method of any of the preceding claims, wherein the disease state is cancer.

11. The method of any of the preceding claims, wherein the electromagnetic radiation of the pump pulse and the stimulant pulse comprise electromagnetic radiation have a frequency in the range of infra-red to ultraviolet light.

12. The method of any of the preceding claims, further comprising producing the pump pulse and the stimulant pulse from a single laser.

13. The method of any of the preceding claims, further comprising producing digital data from a portion of the signal pulse not produced during the exposing of the sample to the stimulant pulse.

14. The method of any of the preceding claims, further comprising producing an image from the digital data.

15. A method of producing an image, comprising collecting a set of data,

wherein the data comprises a plurality of digital data produced by the method of any of the preceding claims.

16. The method of any of the preceding claims, wherein the image is an OCT image.

17. The method of any of the preceding claims, wherein the pump pulse is chirped.

18. The method of any of the preceding claims, wherein the laser has thousands of cm^{-1} of bandwidth.

19. The method of any of the preceding claims, wherein the laser is selected from the group consisting of Ti-sapphire lasers, Cr:forsterite lasers, dye lasers, Yb:YAG lasers, Yb-silica fiber lasers, Er-silica fiber lasers and Nd:glass lasers.

20. In a method of examining a sample by CARS or CSRS, including exposing a sample to laser light to produce a signal, and producing a data set from the signal, the improvement comprising interfering the signal with a reference pulse, and excluding at least a portion of the signal containing non-resonant electromagnetic radiation in producing the data set.

21. The method of any of the preceding claims, comprising:

a step for producing a CARS or CSRS signal pulse of electromagnetic radiation from a sample, and

a step for determining which portions of the signal pulse contain electromagnetic radiation produced by four-wave mixing.

22. A method of examining a sample, comprising:

a step for producing a CARS or CSRS signal pulse of electromagnetic radiation from a sample, and

a step for determining which portions of the signal pulse contain electromagnetic radiation produced by four-wave mixing.

23. An image, produced by the method of any of the preceding claims.

24. An image, produced by the method of any of the preceding claims.

25. The method of any of the preceding claims, wherein the stimulant pulse is chirped.

26. The method of any of the preceding claims, wherein the stimulant pulse comprises Stokes photons.

27. The method of any of the preceding claims, wherein the stimulant pulse comprises anti-Stokes photons.

28. The method of any of the preceding claims, further comprising producing the pump pulse and the stimulant pulse from broadband light produced from bulk materials.

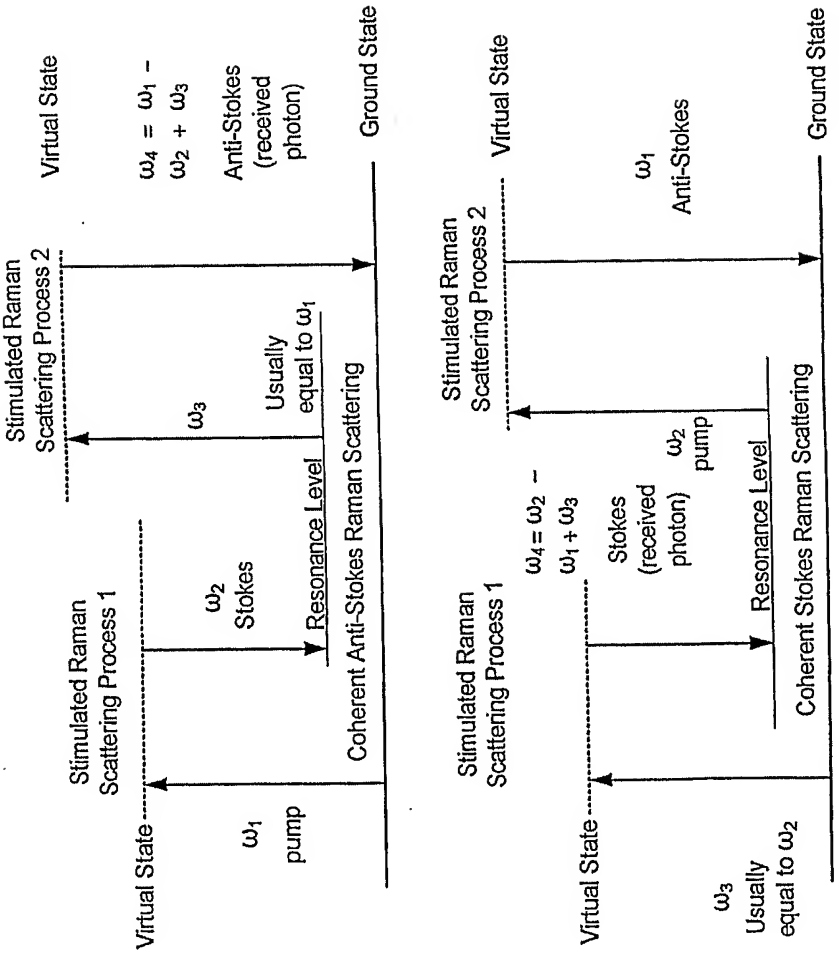

Stimulated Raman Scattering Process 2
Virtual State
$\omega_4 = \omega_1 - \omega_2 + \omega_3$
Anti-Stokes (received photon)
ω_3
Usually equal to ω_1
Resonance Level
Coherent Anti-Stokes Raman Scattering
Ground State
Stimulated Raman Scattering Process 1
ω_2 Stokes
Virtual State
ω_1 pump
Stimulated Raman Scattering Process 2
Virtual State
$\omega_4 = \omega_2 - \omega_1 + \omega_3$
ω_1 Anti-Stokes
Stokes (received photon)
ω_2 pump
Stimulated Raman Scattering Process 1
Resonance Level
Virtual State
ω_3 Usually equal to ω_2
Coherent Stokes Raman Scattering
Ground State

Fig. 1

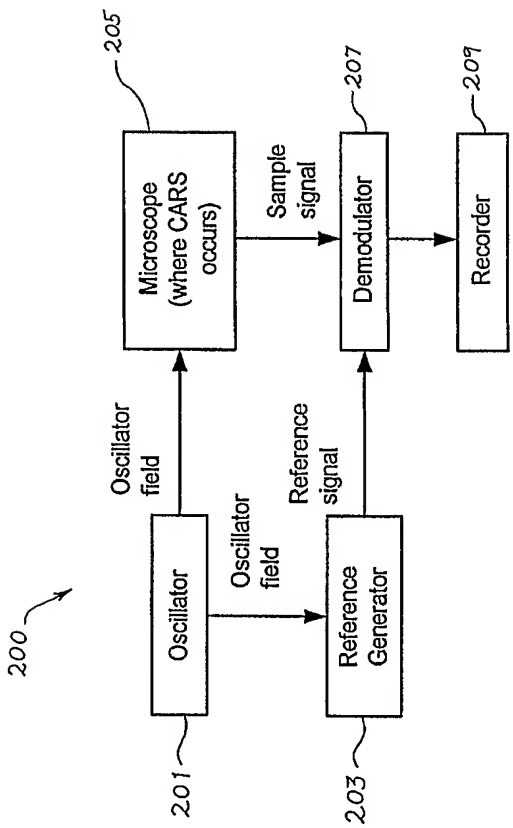

200
Oscillator
201
Oscillator field
Reference Generator
203
Oscillator field
Reference signal
Microscope (where CARS occurs)
205
Sample signal
Demodulator
207
Recorder
209

Fig. 2

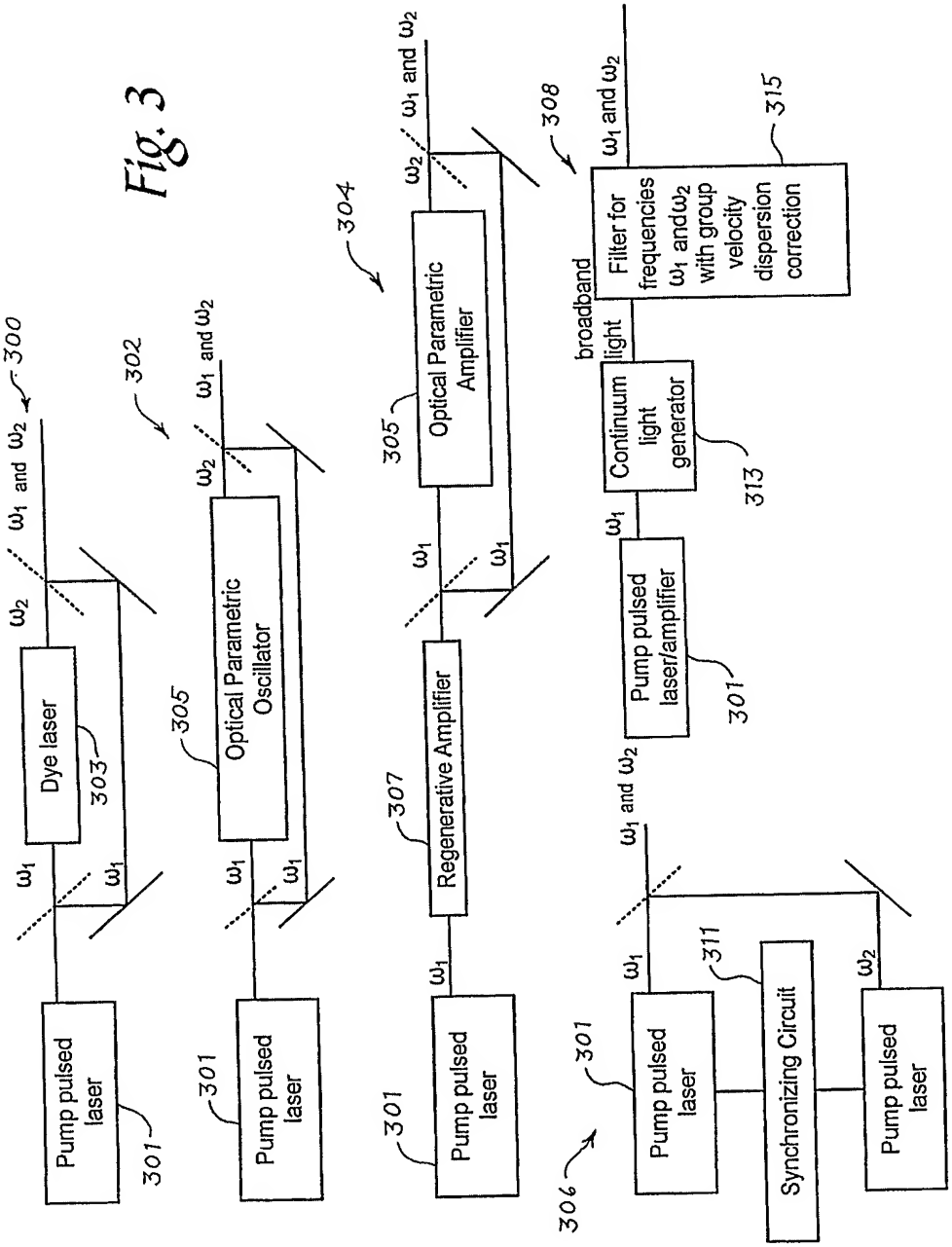

Pump pulsed laser
301
Dye laser
303
ω1
ω2
ω1 and ω2
300
302
Optical Parametric Oscillator
305
304
Optical Parametric Amplifier
Regenerative Amplifier
307
308
broadband light
Continuum light generator
313
Filter for frequencies ω1 and ω2 with group velocity dispersion correction
315
Pump pulsed laser/amplifier
306
Synchronizing Circuit
311

Fig. 3

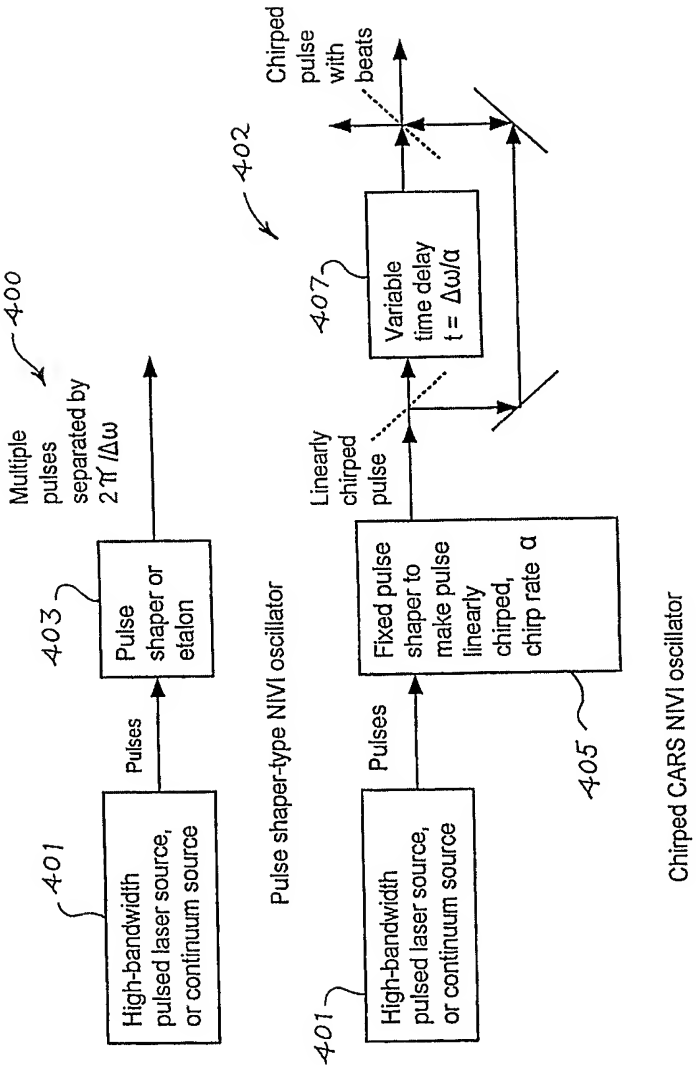

400
Multiple pulses separated by 2π/Δω
401
High-bandwidth pulsed laser source, or continuum source
Pulses
403
Pulse shaper or etalon
Pulse shaper-type NIVI oscillator
402
Chirped pulse with beats
407
Variable time delay t = Δω/α
Linearly chirped pulse
401
High-bandwidth pulsed laser source, or continuum source
Pulses
Fixed pulse shaper to make pulse linearly chirped, chirp rate α
405
Chirped CARS NIVI oscillator

Fig. 4

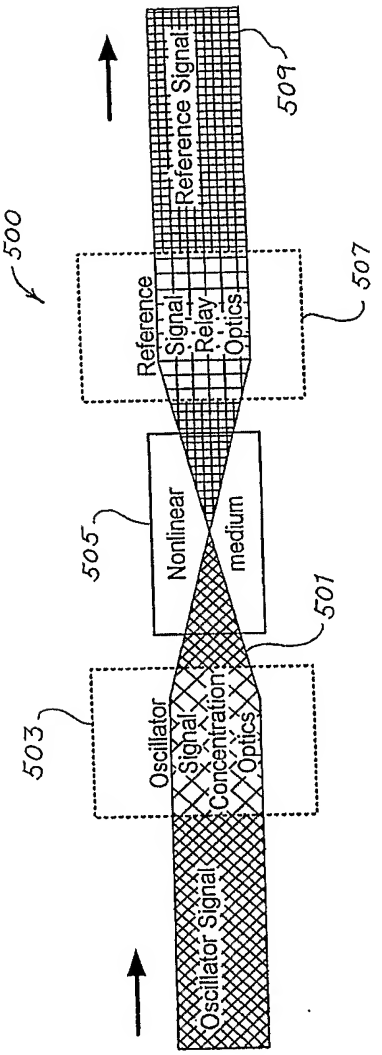

500
509
Reference Signal
507
Reference Signal Relay Optics
505
Nonlinear medium
501
503
Oscillator Signal Concentration Optics
Oscillator Signal

Fig. 5

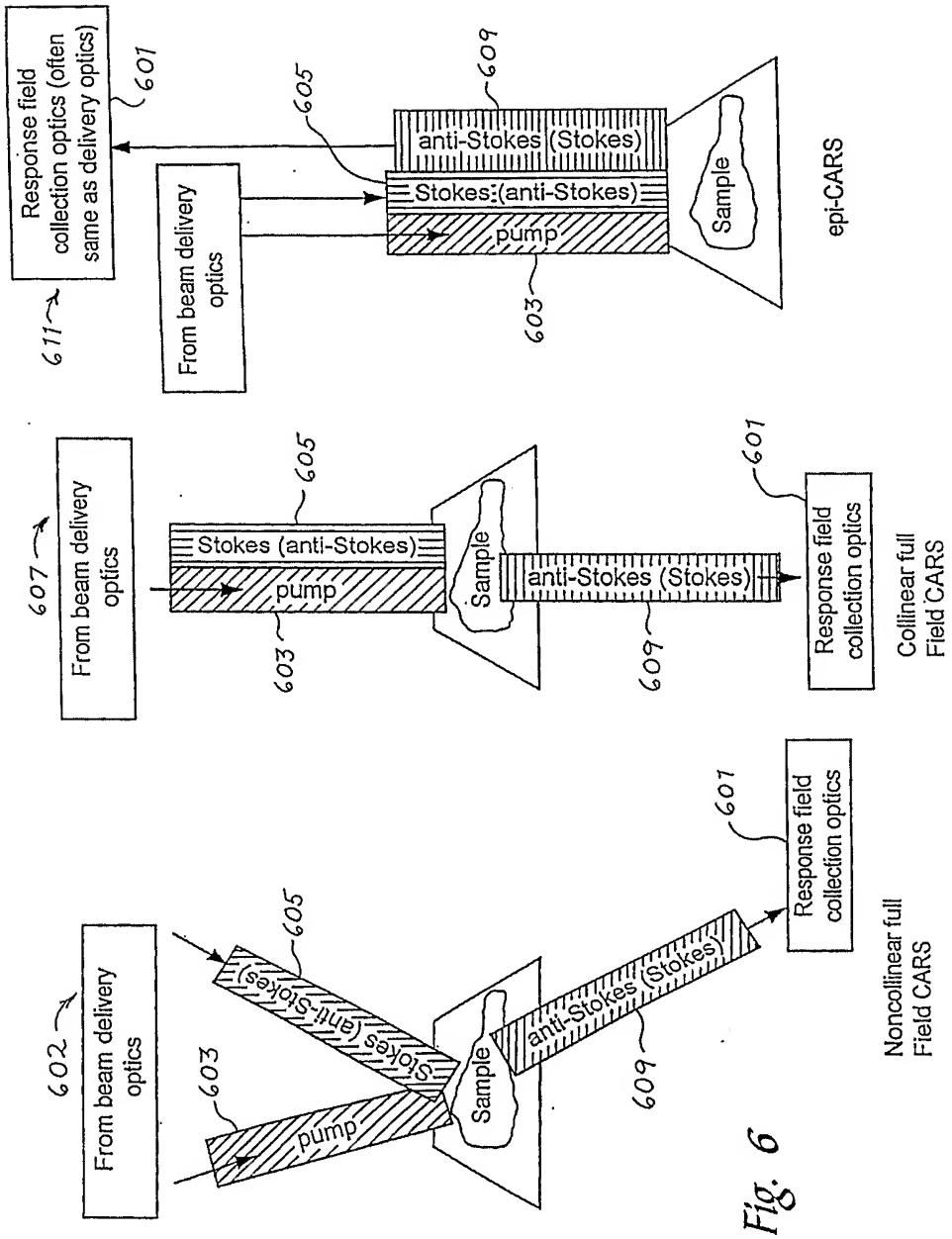

Response field collection optics (often same as delivery optics)
601
611
From beam delivery optics
605
609
anti-Stokes (Stokes)
Stokes (anti-Stokes)
pump
603
Sample
epi-CARS
From beam delivery optics
607
605
Stokes (anti-Stokes)
pump
603
Sample
anti-Stokes (Stokes)
609
601
Response field collection optics
Collinear full Field CARS
From beam delivery optics
602
605
Stokes (anti-Stokes)
603
pump
Sample
anti-Stokes (Stokes)
609
601
Response field collection optics
Noncollinear full Field CARS

Fig. 6

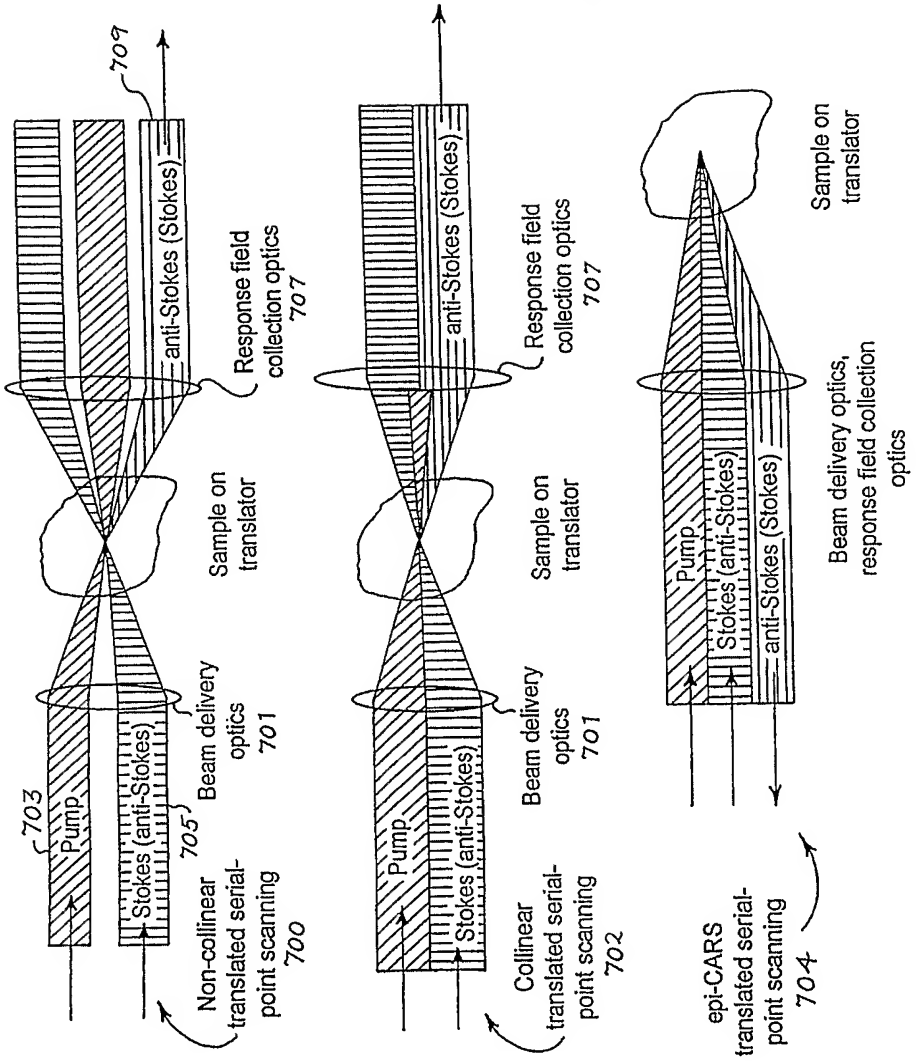

703
Pump
Stokes (anti-Stokes)
705
Beam delivery optics 701
Sample on translator
anti-Stokes (Stokes)
709
Response field collection optics 707
Non-collinear translated serial-point scanning 700
Collinear translated serial-point scanning 702
epi-CARS translated serial-point scanning 704
Beam delivery optics, response field collection optics

Fig. 7

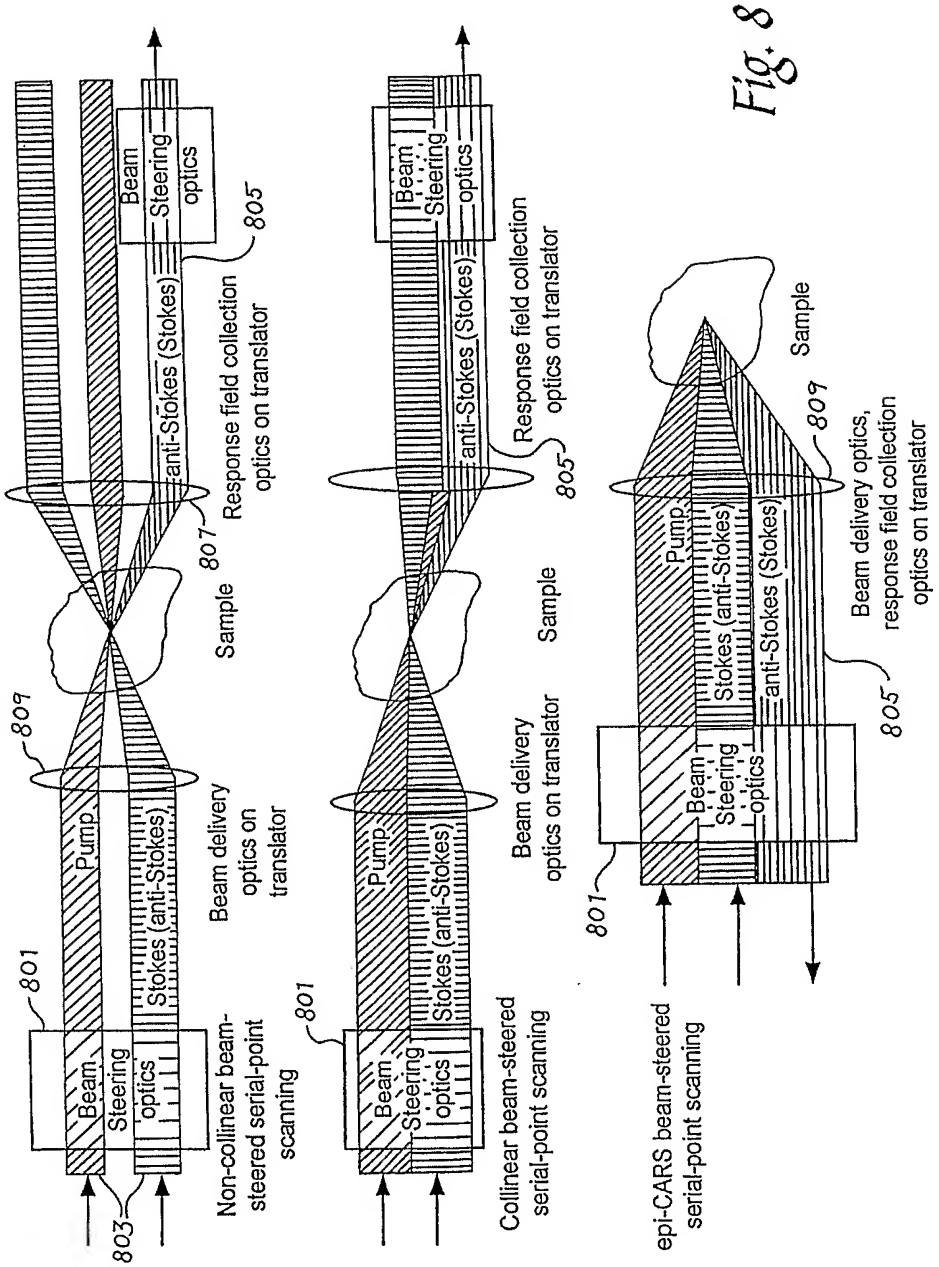

801
Beam
Steering
optics
Pump
Stokes (anti-Stokes)
anti-Stokes (Stokes)
803
809
Beam delivery optics on translator
Sample
807
Response field collection optics on translator
805
Non-collinear beam-steered serial-point scanning
Collinear beam-steered serial-point scanning
epi-CARS beam-steered serial-point scanning
Beam delivery optics, response field collection optics on translator

Fig. 8

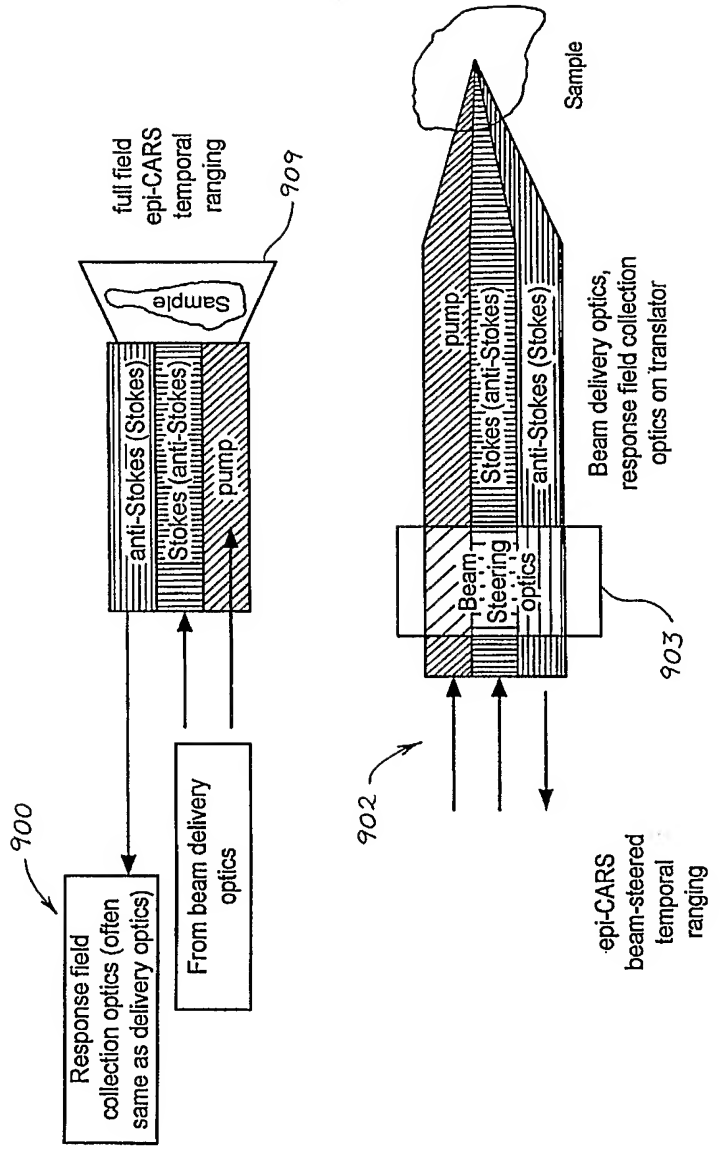

Sample
pump
Stokes (anti-Stokes)
anti-Stokes (Stokes)
Beam Steering optics
Beam delivery optics, response field collection optics on translator
903
902
epi-CARS beam-steered temporal ranging
full field epi-CARS temporal ranging
909
Sample
anti-Stokes (Stokes)
Stokes (anti-Stokes)
pump
From beam delivery optics
900
Response field collection optics (often same as delivery optics)

Fig. 9

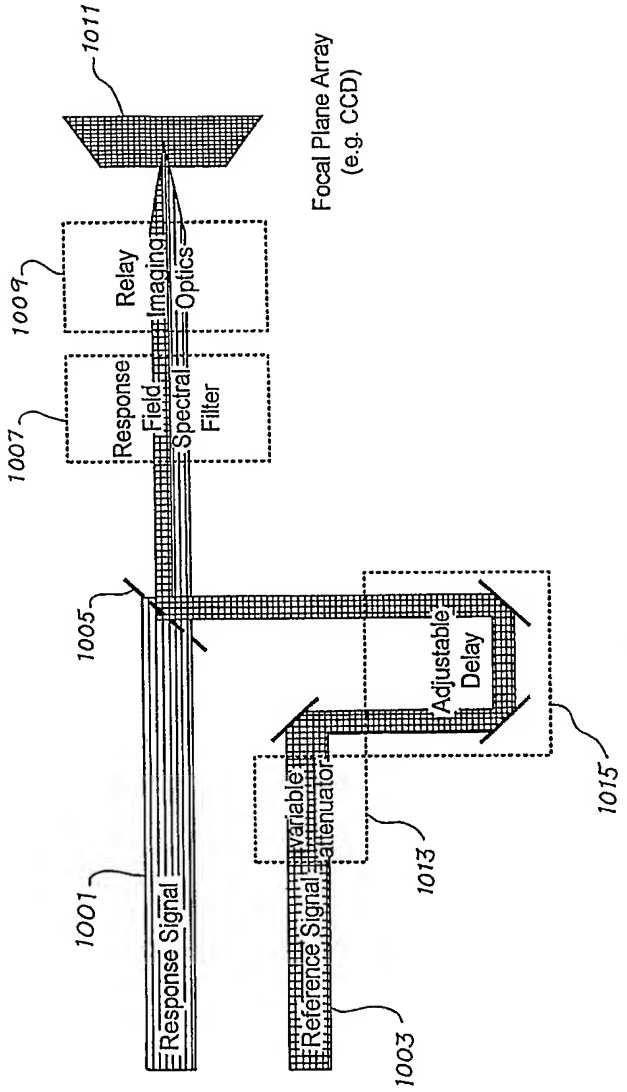

1011
Focal Plane Array
(e.g. CCD)
1009
Relay Imaging Optics
1007
Response Field Spectral Filter
1005
Adjustable Delay
1015
1001
Response Signal
Reference Signal
Variable Attenuator
1013
1003

Fig. 10

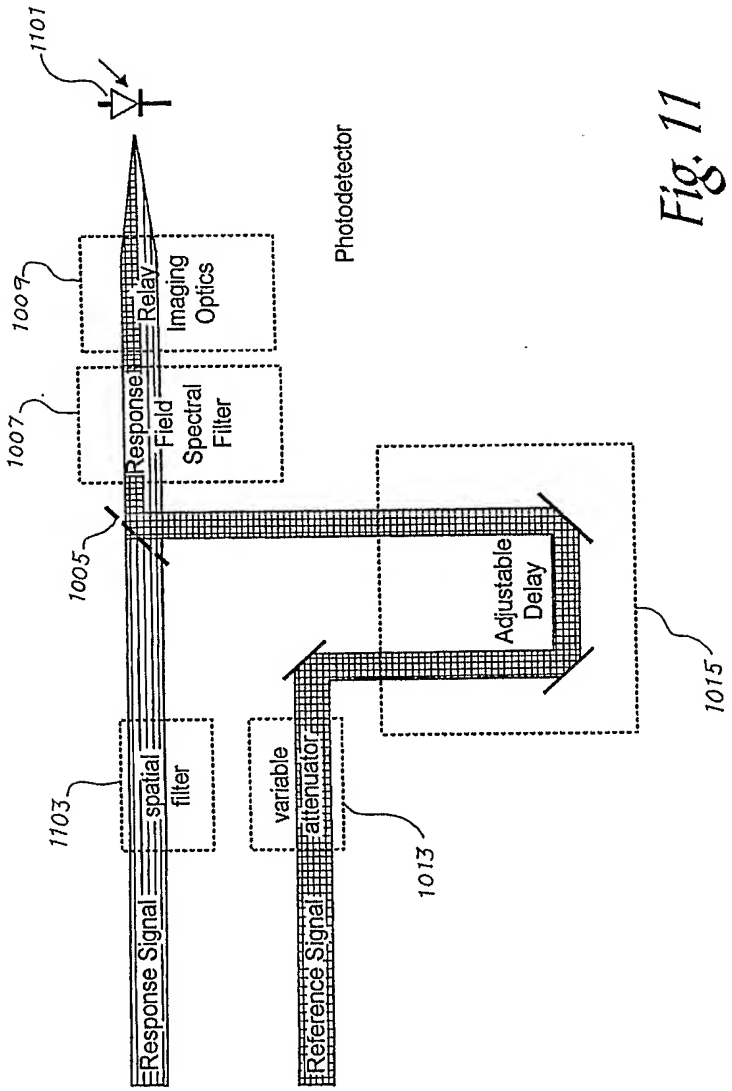

1101
Photodetector
1009
Imaging Optics
Relay
1007
Response Field
Spectral Filter
1005
Adjustable Delay
1015
1103
spatial filter
variable attenuator
1013
Response Signal
Reference Signal

Fig. 11

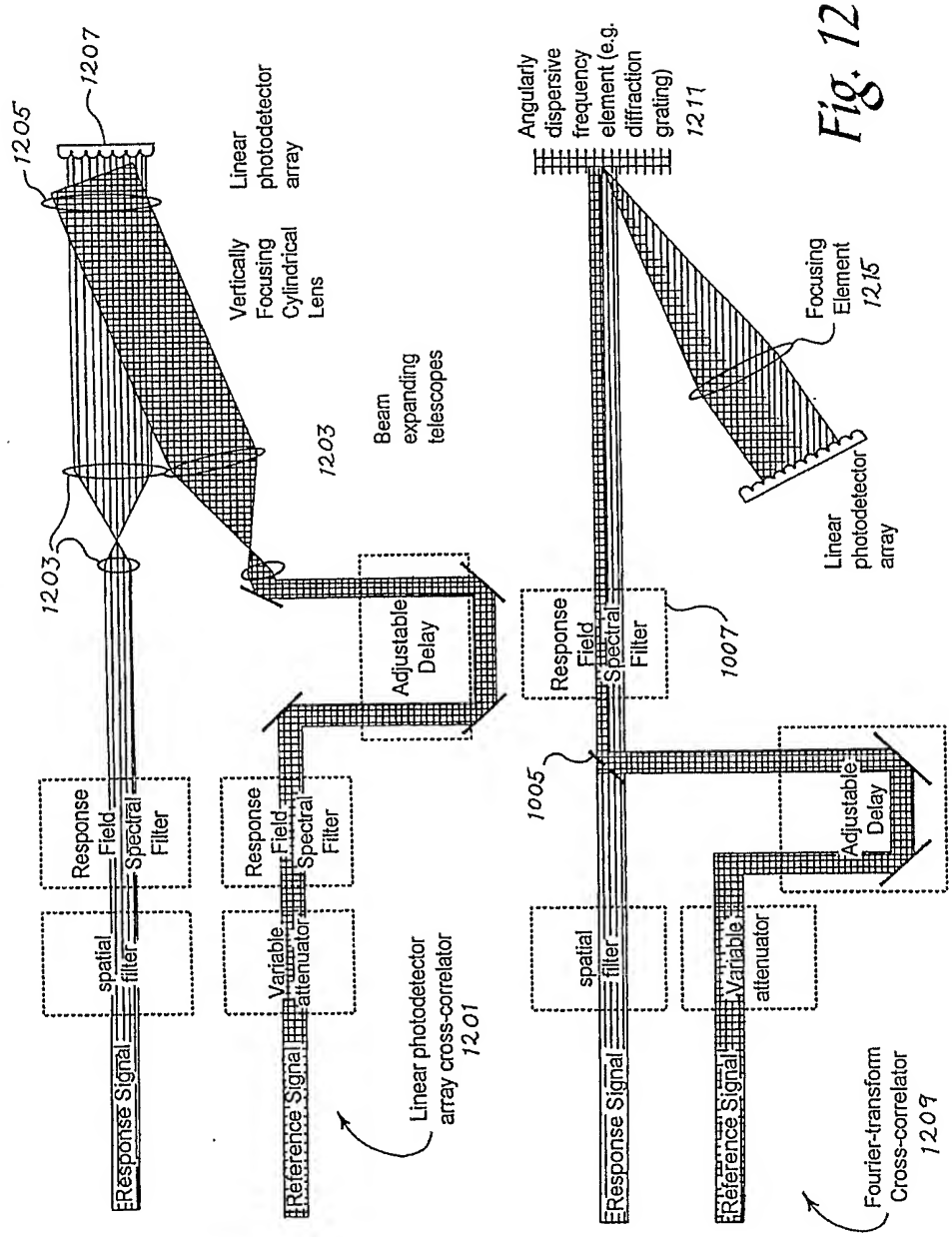

1205
1207
1203
Linear photodetector array
Vertically Focusing Cylindrical Lens
1203
Beam expanding telescopes
Angularly dispersive frequency element (e.g. diffraction grating)
1211
Focusing Element
1215
Linear photodetector array
Response Field Spectral Filter
Adjustable Delay
Response Field Spectral Filter
1007
1005
spatial filter
Variable attenuator
Response Signal
Reference Signal
Linear photodetector array cross-correlator
1201
spatial filter
Variable attenuator
Adjustable Delay
Response Signal
Reference Signal
Fourier-transform Cross-correlator
1209

Fig. 12

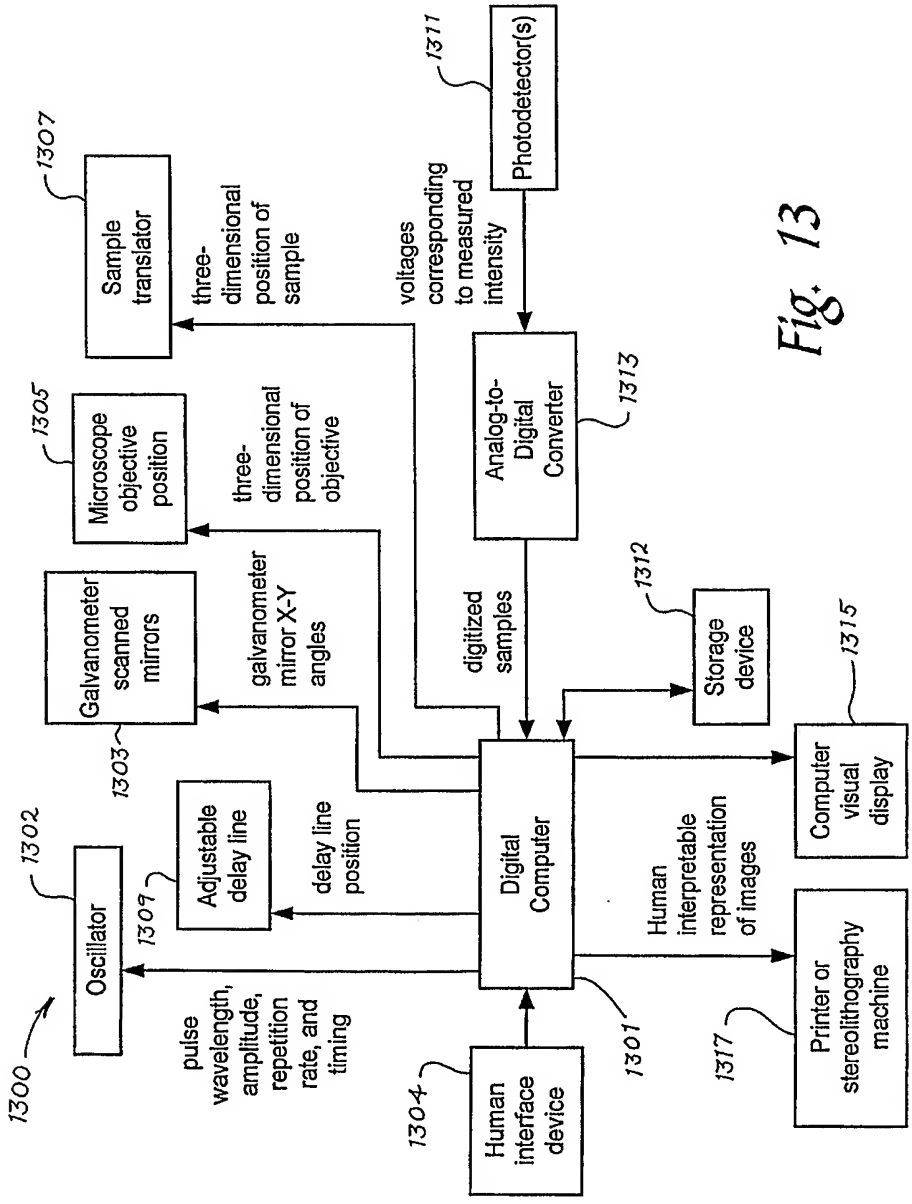

1300
1302
Oscillator
1307
Sample translator
three-dimensional position of sample
1305
Microscope objective position
three-dimensional position of objective
1303
Galvanometer scanned mirrors
galvanometer mirror X-Y angles
1309
Adjustable delay line
delay line position
pulse wavelength, amplitude, repetition rate, and timing
1304
Human interface device
1301
Digital Computer
1311
Photodetector(s)
voltages corresponding to measured intensity
1313
Analog-to-Digital Converter
digitized samples
1312
Storage device
1315
Computer visual display
Human interpretable representation of images
1317
Printer or stereolithography machine

Fig. 13

Fig. 14

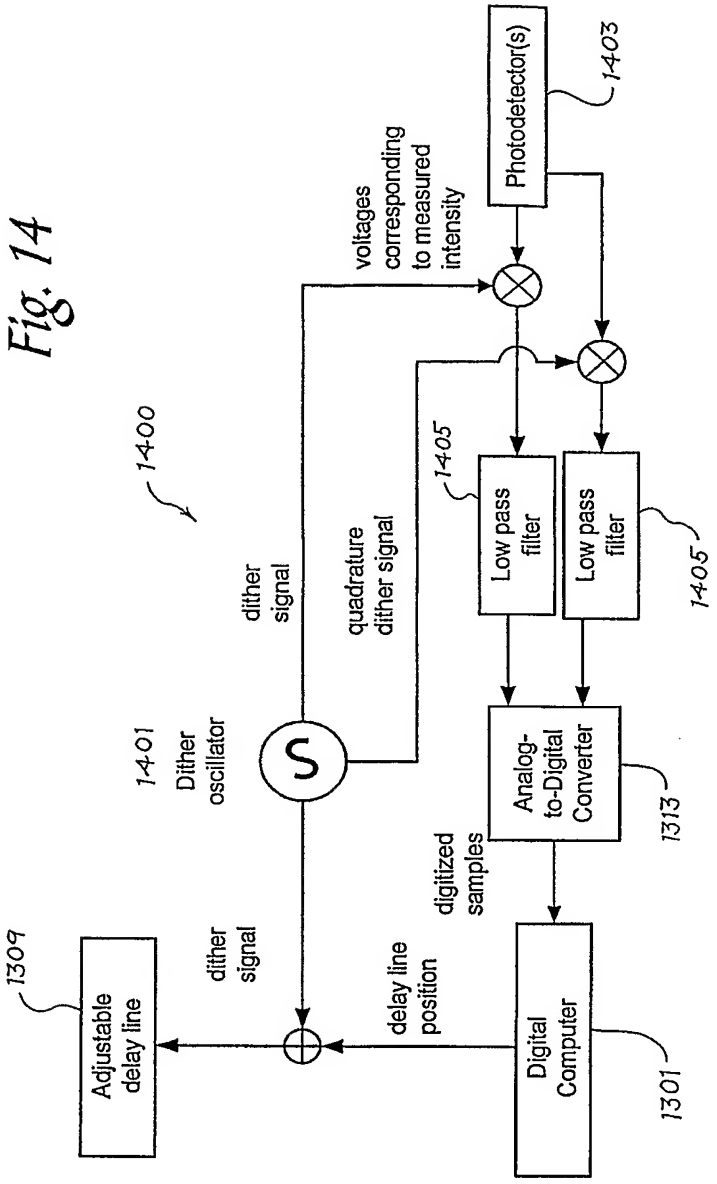

1400
1401
Dither oscillator
S
dither signal
quadrature dither signal
voltages corresponding to measured intensity
Photodetector(s)
1403
1405
Low pass filter
Low pass filter
1405
Analog-to-Digital Converter
1313
digitized samples
Digital Computer
1301
delay line position
dither signal
1309
Adjustable delay line

Figure 15

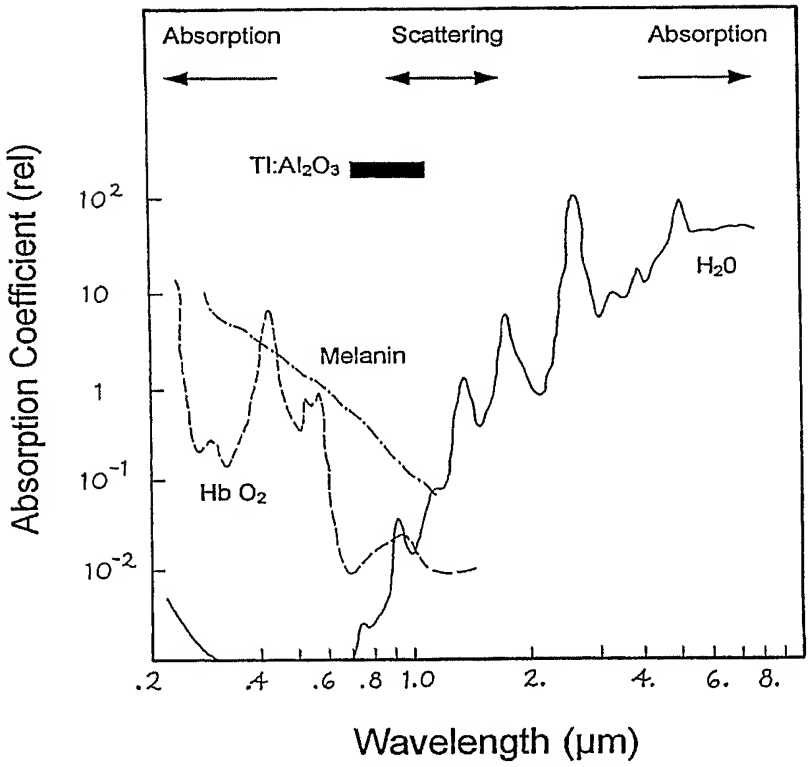

Absorption
Scattering
Absorption
Tl:Al2O3
H20
Melanin
Hb O2
Absorption Coefficient (rel)
10^2
10
1
10^-1
10^-2
.2
.4
.6
.8
1.0
2.
4.
6.
8.
Wavelength (μm)

Figure 16

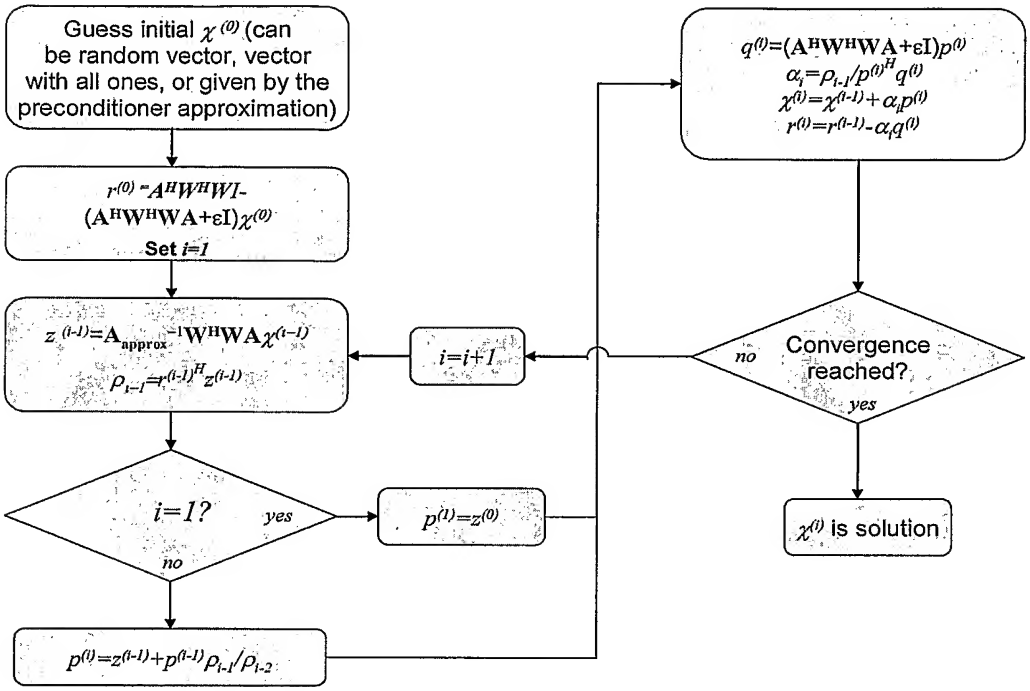

Guess initial $\chi^{(0)}$ (can be random vector, vector with all ones, or given by the preconditioner approximation)
$r^{(0)}=A^HW^HWI-(A^HW^HWA+\varepsilon I)\chi^{(0)}$
Set $i=1$
$z^{(i-1)}=A_{approx}^{-1}W^HWA\chi^{(i-1)}$
$\rho_{i-1}=r^{(i-1)H}z^{(i-1)}$
$i=i+1$
$q^{(i)}=(A^HW^HWA+\varepsilon I)p^{(i)}$
$\alpha_i=\rho_{i-1}/p^{(i)H}q^{(i)}$
$\chi^{(i)}=\chi^{(i-1)}+\alpha_i p^{(i)}$
$r^{(i)}=r^{(i-1)}-\alpha_i q^{(i)}$
no
Convergence reached?
yes
$\chi^{(i)}$ is solution
$i=1?$
yes
no
$p^{(1)}=z^{(0)}$
$p^{(i)}=z^{(i-1)}+p^{(i-1)}\rho_{i-1}/\rho_{i-2}$

Figure 17

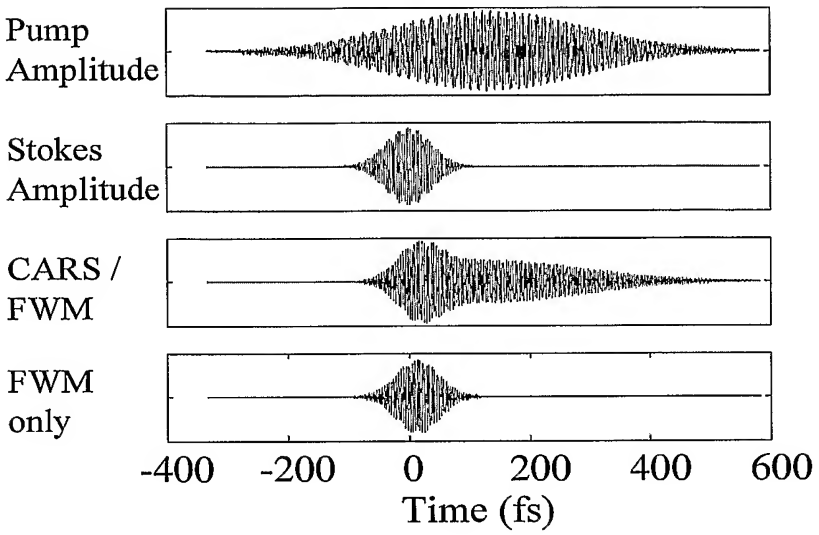

Pump Amplitude
Stokes Amplitude
CARS / FWM
FWM only
-400
-200
0
200
400
600
Time (fs)

Figure 18

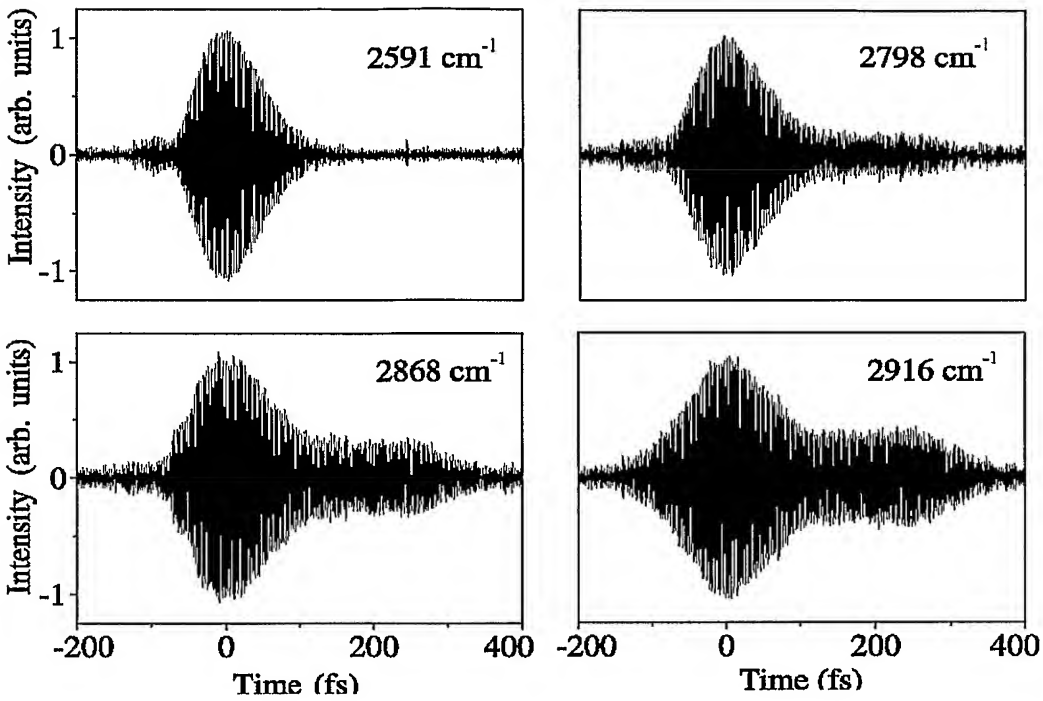

2591 cm-1
2798 cm-1
2868 cm-1
2916 cm-1
Intensity (arb. units)
Intensity (arb. units)
1
0
-1
-200
0
200
400
Time (fs)
Time (fs)

Figure 19

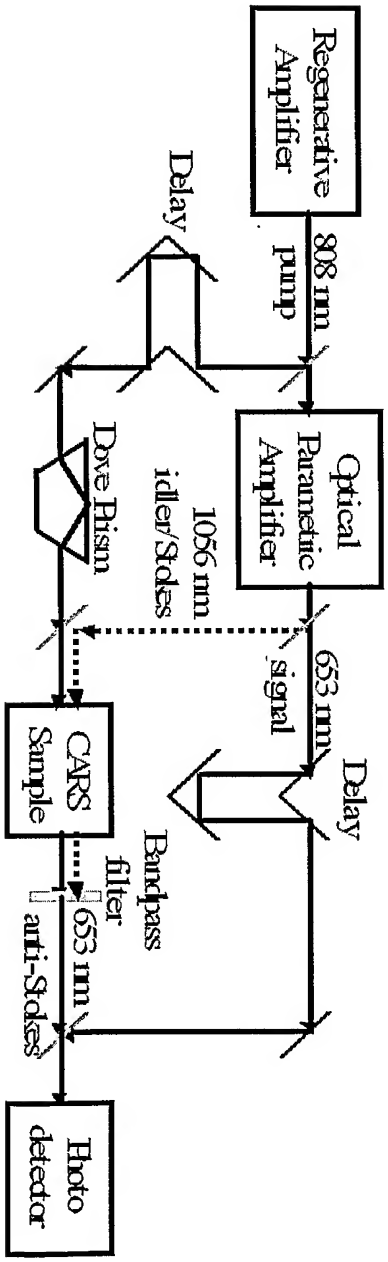

Regenerative Amplifier
808 nm pump
Delay
Optical Parametric Amplifier
1056 nm idler/Stokes
Dove Prism
653 nm signal
Delay
CARS Sample
Bandpass filter
653 nm anti-Stokes
Photo detector

Figure 20

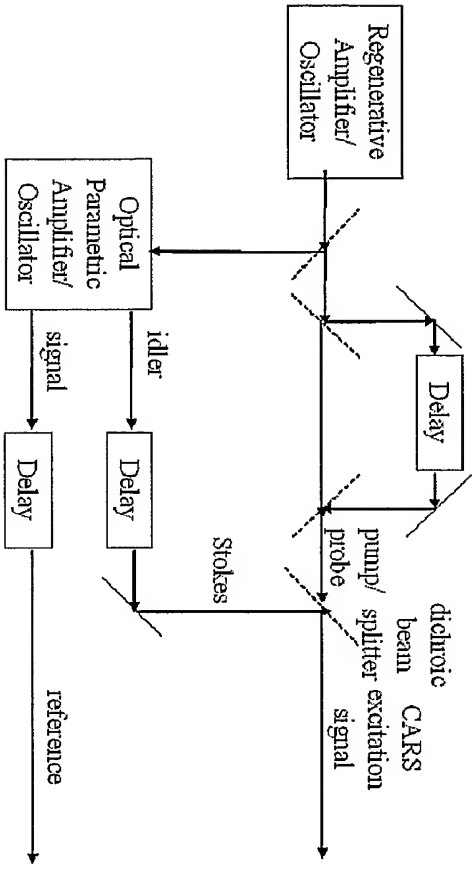

Regenerative Amplifier/ Oscillator
Optical Parametric Amplifier/ Oscillator
signal
idler
Delay
Delay
Delay
Delay
Stokes
pump/ probe
dichroic beam splitter
CARS excitation signal
reference

Figure 21

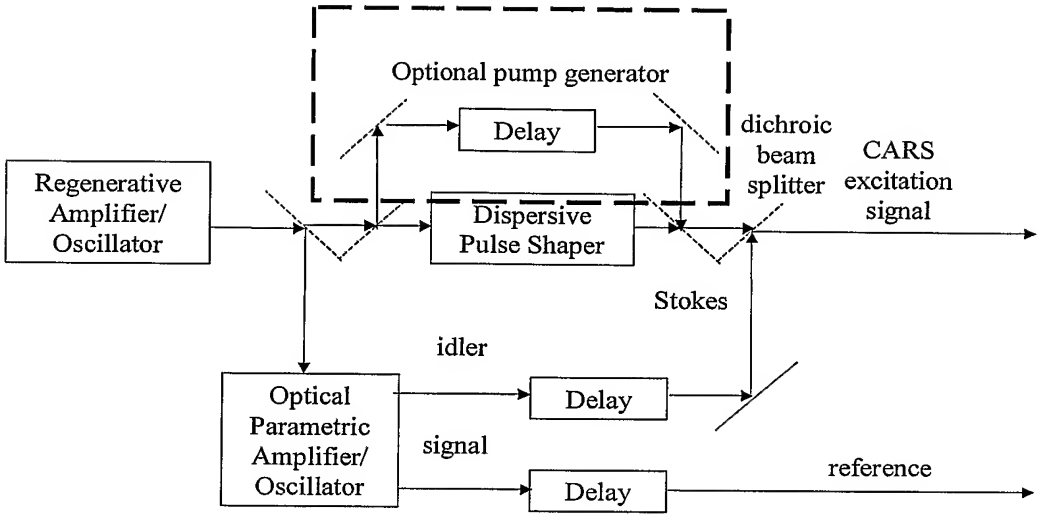

Optional pump generator
Delay
dichroic beam splitter
CARS excitation signal
Regenerative Amplifier/ Oscillator
Dispersive Pulse Shaper
Stokes
idler
Optical Parametric Amplifier/ Oscillator
Delay
signal
reference
Delay

Figure 22

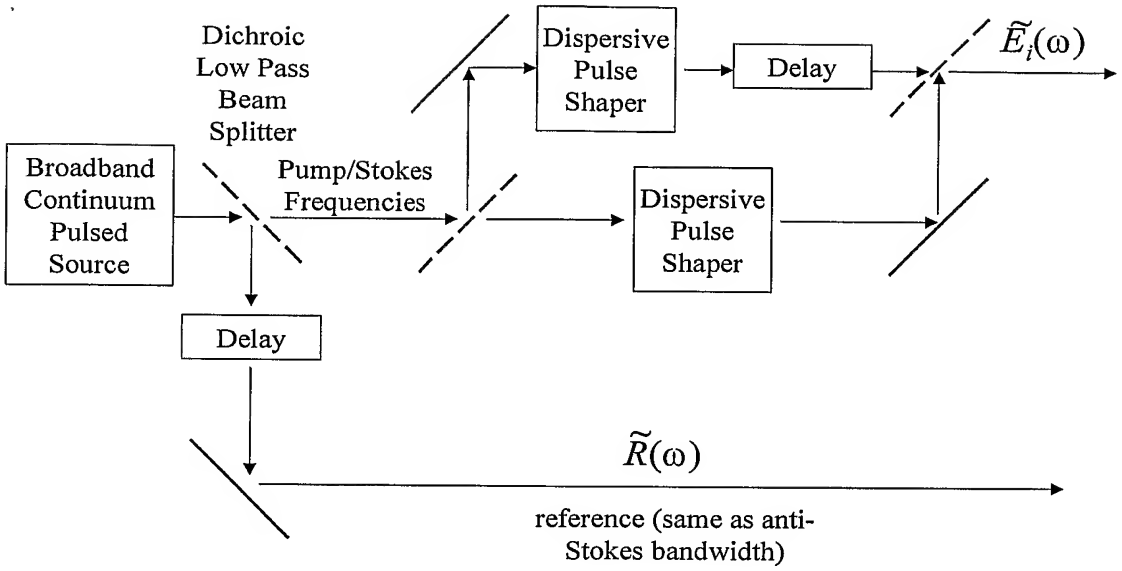

Dichroic
Low Pass
Beam
Splitter
Broadband
Continuum
Pulsed
Source
Pump/Stokes
Frequencies
Dispersive
Pulse
Shaper
Delay
$\tilde{E}_i(\omega)$
Dispersive
Pulse
Shaper
Delay
$\tilde{R}(\omega)$
reference (same as anti-
Stokes bandwidth)

Figure 23

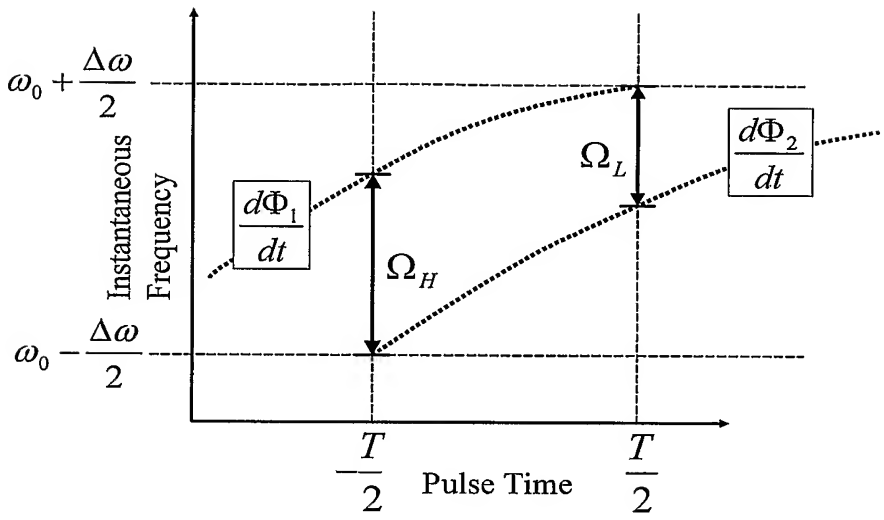

$\omega_0 + \frac{\Delta\omega}{2}$
$\omega_0 - \frac{\Delta\omega}{2}$
Instantaneous Frequency
$\frac{d\Phi_1}{dt}$
$\frac{d\Phi_2}{dt}$
Ω_H
Ω_L
$-\frac{T}{2}$
Pulse Time
$\frac{T}{2}$

Figure 24

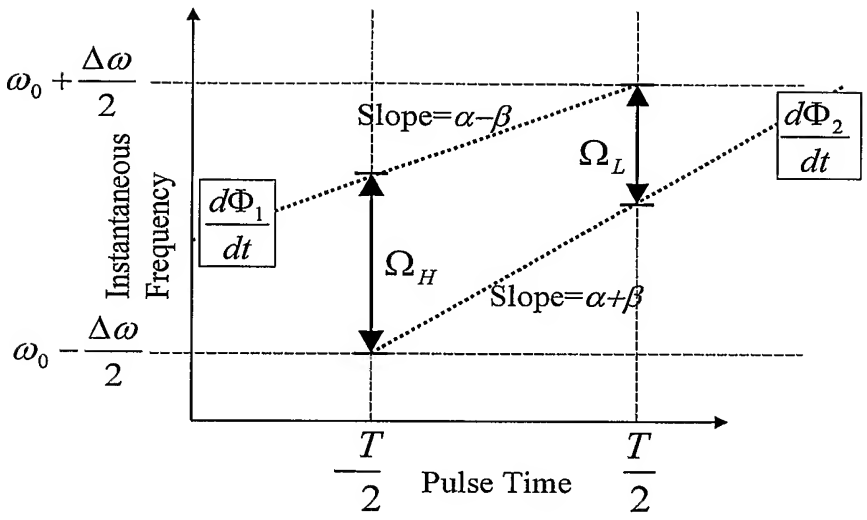

$\omega_0 + \frac{\Delta\omega}{2}$
$\omega_0 - \frac{\Delta\omega}{2}$
Instantaneous Frequency
Slope=$\alpha-\beta$
Slope=$\alpha+\beta$
$\frac{d\Phi_1}{dt}$
$\frac{d\Phi_2}{dt}$
Ω_L
Ω_H
$\frac{T}{2}$
Pulse Time
$\frac{T}{2}$

Figure 25

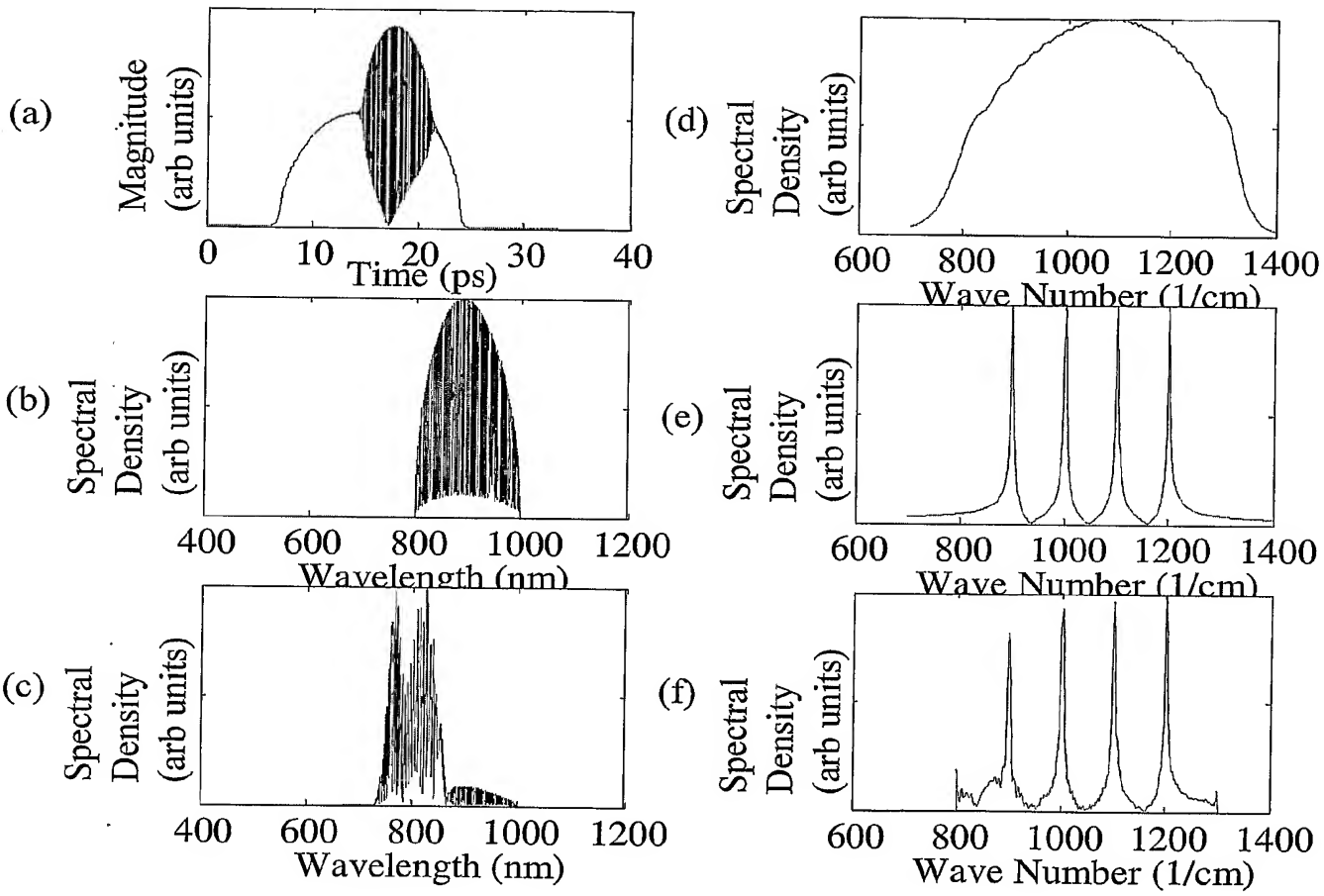
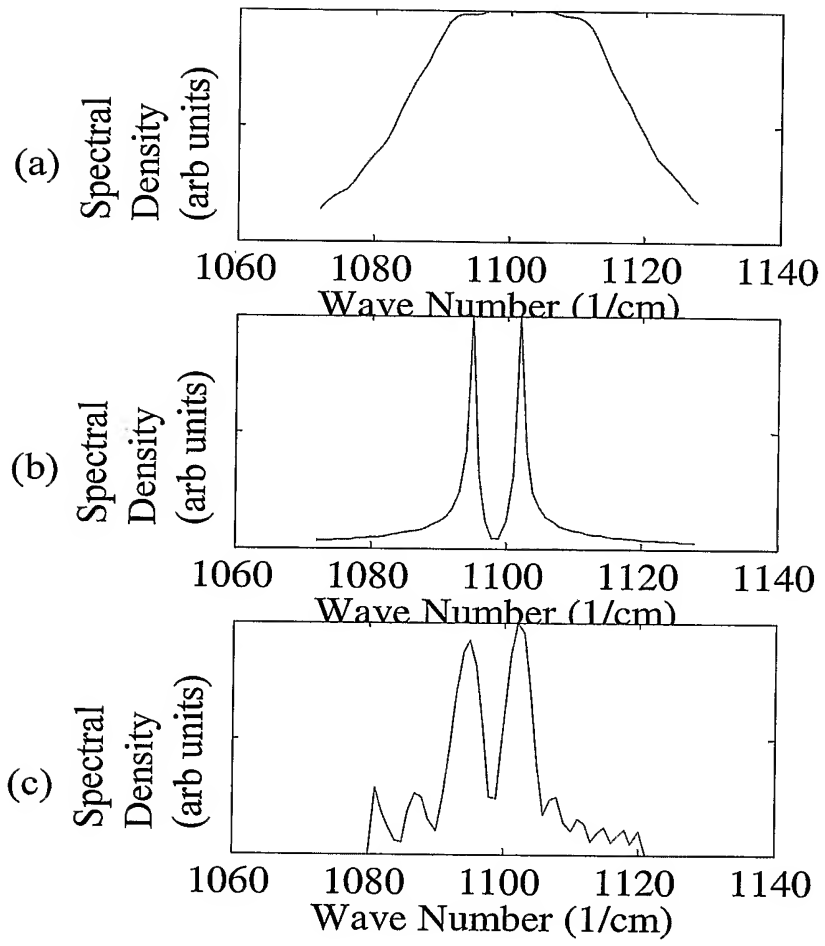

(a)
Magnitude (arb units)
0
10
20
30
40
Time (ps)
(b)
Spectral Density (arb units)
400
600
800
1000
1200
Wavelength (nm)
(c)
Spectral Density (arb units)
400
600
800
1000
1200
Wavelength (nm)
(d)
Spectral Density (arb units)
600
800
1000
1200
1400
Wave Number (1/cm)
(e)
Spectral Density (arb units)
600
800
1000
1200
1400
Wave Number (1/cm)
(f)
Spectral Density (arb units)
600
800
1000
1200
1400
Wave Number (1/cm)

Figure 26


(a)
Spectral Density (arb units)
1060
1080
1100
1120
1140
Wave Number (1/cm)
(b)
Spectral Density (arb units)
1060
1080
1100
1120
1140
Wave Number (1/cm)
(c)
Spectral Density (arb units)
1060
1080
1100
1120
1140
Wave Number (1/cm)